Title
Essays in Industrial Organization

Author
Yarmosh, Yevheniia Hryhorivna

Publication Date
2019

Peer reviewed|Thesis/dissertation

Essays in Industrial Organization

by

Yevheniia Hryhorivna Yarmosh

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Handel, Chair
Professor Kei Kawai
Professor Giovanni Compiani

Summer 2019

Essays in Industrial Organization

Abstract

Essays in Industrial Organization

by

Yevheniia Hryhorivna Yarmosh

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Benjamin Handel, Chair

This dissertation consists of two essays in the fields of Industrial Organization and Behavioral Economics. It studies the issue of consumers' biased beliefs that result in misestimation of personal consumption and excessive spending on the products that will never be used in the future.

Chapter 1 introduces the issue of overspending in the videogames market. The excessive purchasing exhibits itself in the behavior when consumers purchase games and never open them later. According to the descriptive statistics, around 29% of games in people's game libraries are unopened as of March 18, 2019. This setting is very beneficial for the analysis in question as besides purchasing decisions one can observe consumption patterns as well, which are usually problematic to track. I gather the main dataset from Steam, one of the biggest online gaming platforms. The data contain daily purchasing and consumption statistics as well as daily prices in all supported currencies for the period of 9 months. In this chapter I study economic and behavioral mechanisms that may be causing such conduct. In particular, I perform survival analysis, discuss search costs, study the effects of game heterogeneity, check for the presence of the projection bias, and consider intertemporal substitution of purchases as possible explanations of the puzzle. I show that none of the above mentioned mechanisms are supported by the dataset and arrive to the conclusion that the main reason for the excessive spending is consumers' biased beliefs regarding their future willingness to try new products.

Chapter 2 builds on the results of the first chapter. In this part of the thesis I develop a structural model that takes into account people's biased beliefs at the point of purchasing new products. It is a static model with the elements of dynamic behavior where every period consumers decide how many games they want to purchase based on how many games they stopped playing in the previous period and then decide how much time to devote to playing. I discuss functional and distributional assumptions that I impose in order to identify the parameters of the model. Next I simulate the model and run three counterfactuals that could potentially alleviate the ovespending behavior. In particular, I consider debiasing consumers, removing sales, and decreasing variety (i.e. imposing a lower bound on game

quality). According to the simulated results, removing sales only slightly decreases over-spending. Decreasing variety also reduces excessive spending, but it is the most effective for low userscore thresholds. One has to determine the optimal threshold that minimizes the effect of the bias. Debiasing consumers is the most effective in the reduction of expenditures, and the total effect depends on the initial magnitude of the bias.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express my deepest gratitute to my thesis advisor Benjamin Handel for his guidance and advice that he has provided over the course of my graduate studies. I am also very thankful to Kei Kawai for his time, attention and constructive feedback that helped me improve my research and presentation skills. I greatly appreciate the honest and extremely detailed suggestions that Dmitry Taubinsky gave me during our meetings. I thank Giovanni Compiani for his thorough comments and recommendations. I am grateful to my oral exam committee members Miguel Villas-Boas, Michael Katz, and Joseph Farrell for their advice and feedback. I would also like to thank Przemyslaw Jeziorski, Minjung Park, Steven Tadelis, and Ned Augenblick for finding the time in their busy schedules to meet with me and provide their suggestions on my research.

My special thanks are extended to Patrick Allen, Vicky Lee, Heather Iwata, and Joe Sibol for their help with any issues or concerns that I had. I appreciate the friendliness of my fellow graduate students and thank them for creating a supportive research environment. Thank you, Tiffany Tsai, for the feedback on my research, your readiness to help and chat with me anytime, and just for being a great friend! I hope that we will meet again sometime. Thank you, Paolo Zacchia, for your hospitality, keen sense of humor and inexhaustible willingness to listen to my troubles and concerns.

The last but not the least, I am indebted to my family for their unconditional love and support. Thank you, Bogdan, for always cheering me up. This has been a long and extremely exhausting journey at times, and I don't think that I would have made it without you being by my side.

Chapter 1

Studying Overspending in the Videogames Market

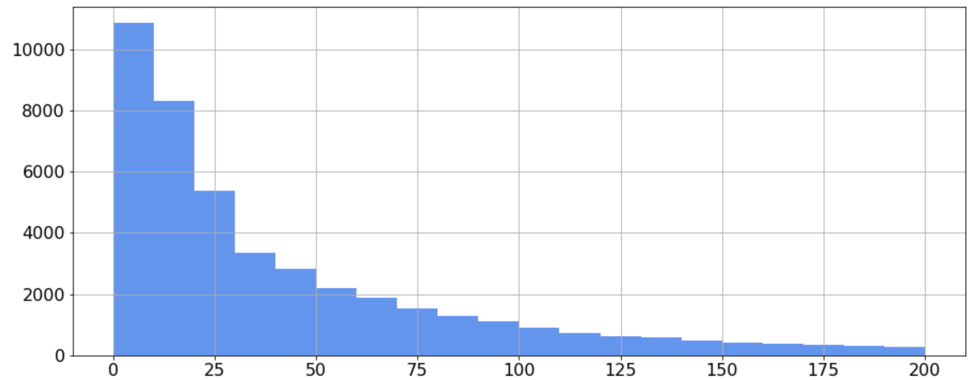
1.1 Introduction

Misestimation of future consumption is a common phenomen in many markets; the most representative examples from the literature would be naive beliefs about future gym attendance (DellaVigna and Malmendier, 2006), suboptimal choice of a cell-phone plan (Grubb and Osborne, 2015), *et cetera*. This paper, however, studies misestimation of future consumption in the context of the videogames market using the setting of Steam online gaming platform. The environment is very convenient as it allows to not only observe individual purchases but consumption as well, which is usually challenging or impossible to track in other settings. The main objective of this research is to study why consumers buy and not even open some of the purchased videogames. According to the sample estimates, on average, around 28.8% of the titles in people's game libraries remain unopened as of March 18, 2019.

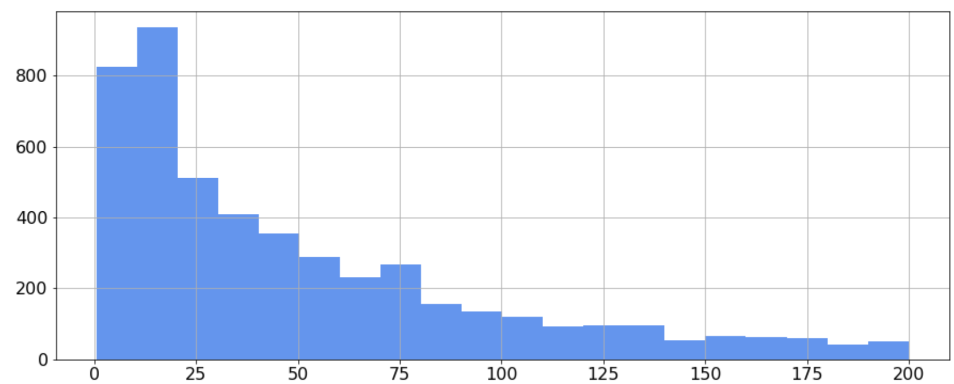
The videogames industry is rather sizeable: according to Gartner (2013), the market was valued at 93 billion dollars in 2013 and was expected to expand to 111 billion dollars in 2015. However, while being large, the industry is not transparent in terms of sales; obtaining the sales data is exceptionally hard. As a general rule, game developers do not publicly report this information. Occasionally, the firms may announce some data, especially when they have reached a certain milestone on having sold a specific game title, and consequently promote the game even more.

Over the past years the game publishers have been shifting the distribution of their products to online platforms due to convenience and reduction of marginal costs; one of such gaming platforms is Steam. Owned by Valve Corporation, it is the largest online distribution platform for PC games. According to the estimates, Steam generated 4.3 billion dollars of game sales in 2017 (PCGamesN, 2017).

A couple of facts serve as motivation for this research: people spend a lot of money on videogames, and approximately half of that money is left on the table as a result of not



(a) All Consumers



(b) US Consumers

Figure 1.1: Distribution of the Money Spent on Unopened Games

Note: the estimates are evaluated in US dollars for the 6.5-months period (June 8 — December 24, 2018), allowing at least 12 weeks to open each game title.

consuming all purchases. I use my sample to estimate that individuals have spent \$117 per person on videogames (\$170 per person conditional on making a purchase) over the 9-months period; US consumers[1], in particular, have spent \$250 per person (\$308 per capita conditional on a purchase) over the same time period. As for the second part of the statement, I find that all sample consumers have spent \$12.8 mln on games, 50.1% of which (\$6.4 mln) went towards games that are still unopened. The statistics across US consumers is very similar: people have spent \$2.6 mln on games, and 52.9% (\$1.4 mln) was spent on unopened game titles. Figure 1.1 shows the distribution of the money allocated to unopened games over 6.5 months. The mean is approximately \$58 per person (\$113 for US consumers), which roughly equals \$107 per person (\$209 for US consumers) per year.

In order to explain the overspending puzzle I scrape daily data on individuals' playing

[1]US consumers constitute approximately 10% of the sample

statistics, game prices in all currencies supported by the platform, and general information about game titles. The main advantage of my dataset is the ability to track daily consumption and purchases on the micro level. The panel consists of approximately 105,000 individuals. Most of them are active shoppers, but not many opened all of their recent purchases: 77% of consumers made at least one purchase during the 9-months period, but only 40.2% of active purchasers opened every paid game that they had bought. 45.8% of sample consumers have at lease one unopened recently purchased game.

The rest of the chapter proceeds as follows. Section 1.2 covers literature review. Section 1.3 describes the dataset in more details and provides descriptive statistics. I show that consumers stop playing less games than they buy over the same time period while keeping the number of distinct games played fairly constant. Consequently, people don't have time to open all of the new games before they buy even more games, which results in the stockpiling of unopened games as the time goes by. In Section 1.4 I discuss other mechanisms that may be causing overspending behavior such as presence of transaction costs, heterogeneity in game characteristics, intertemporal substitution of purchases, *et cetera.* I also perform the survival analysis and check for the presence of the projection bias. I find that none of the mechanisms mentioned above are supported by my data and conclude that the most probable reason for the excessive purchasing patterns is individual heterogeneous biased beliefs about wanting to try new games. Section 1.5 draws conclusions.

1.2 Literature Review

This research is related to several bodies of literature: misestimation of consumption, video games market research, and, specifically, Steam platform research.

Consumption Misestimation Research

Generally, overspending is related to the vast literature on consumer choice misestimation and biased beliefs. According to DellaVigna (2009), there are three main forms in which consumers' behavior may differ from the standard models of consumption: nonstandard preferences, nonstandard beliefs, and nonstandard decision making.

The first group of research typically formalizes nonstandard preferences with (β, δ) discounting. Some of this literature also involves overconfidence or naive agents: DellaVigna and Malmendier (2006) incorporate partial naivette to study why consumers overpay for gym membership. Ausubel (1999) studies credit credit card borrowing and finds that individuals overrespond to the initial promotional interest rate, which is consistent with naive beliefs and hence underestimation of the future borrowing. Grubb and Osborne (2015) formalize myopic behavior as underestimation of variance of the future consumption of cell phone services. The puzzle of overspending that I have at hand, however, is conceptually different and thus cannot be explained with (β, δ) discounting even with the addition of full or partial naivete as it involves consumption of immediately pleasurable goods with no deferred cost.

The other group of behavioral research that is closely related to the issue studies projection bias. It is defined as a bias in evaluating future utility based on the current state of an individual (Loewenstein et al., 2003). For example, one may want to order more cold weather apparel when it rains (Conlin et al., 2007), or make future food choices that are biased by the current state of hunger (Read and Van Leeuwen, 1998). However, while this type of studies concerns with bias in utility, the bias studied in this paper is of a different nature.

Videogames Industry Research

The majority of videogames market related research that does not directly study Steam platform examines effects of various factors on demand for videogames, or concentrates on two-sided connection of software and hardware markets. In contrast to this paper, only a few studies from this group examine consumer behavior on the micro level. Standard data come from the NPD Group market research company. Below is a brief overview of the papers from this section of research.

Effect on sales. Gil and Warzynski (2014) study how vertical integration is related to game performance. They find that independent games have lower prices than integrated ones, and this happens because the latter ones have better timing release strategies and game quality overall. Bounie et al. (2008) use an online survey to show that peer reviews as well as expert reviews positively affect consumer demand. Zhu and Zhang (2006) find significant effect of online reviews on sales using differences-in-differences approach.

Software versus hardware, network interactions. Gretz (2010b) structurally estimates a discrete choice demand model and concludes that quality of the hardware of a game console affects demand 11.4 times more than the network size. Venkatraman and Lee (2004) also look at the link between software developers and game consoles using multiprobability regression and find that the following four factors are critical for the network dynamics: density overlap, embeddedness, newness, and dominance. Derdenger (2014) studies how making software that is incompatible with other gaming consoles affects competition in the industry. He concludes that tying increases competition and it should be used by companies with low cost of developing games. Another paper by Gretz (2010a) studies the connection between console prices and software availability. The main result is that less software makes hardware (i.e consoles) less valuable, which is consistent with the theory on two-sided markets.

Other. Bauckhage et al. (2012) use telemetry data to study how fast people lose interest in playing games and find that playtime follows Weibull distribution. There are also economic geography papers that study local externalities from clusters of firms and dynamics of the interfirm networks (De Vaan et al., 2012; Balland et al., 2012).

Steam Platform Research

The literature from this group mainly focuses on estimating demand for videogames, exploring the effects of various factors on sales, and cluster analysis. Due to the difficulties of the microlevel data collection process there is little economic research in this field, and sometimes authors have to seek a compromise between the research question in hand and the frequency of data. Often, if one requires the sales data she has to rely on `steamspy.com` sales estimates. What follows is a short review of the recent papers in this group.

Cluster analysis. Sifa et al. (2015) perform cluster analysis on a 6 million users sample scraped from Steam platform. They mainly focus on individuals' playtime and cross-game behavior. The authors find that most players are mainly dedicated to one game while around one third of players prefer combining multiple games. Another finding is that the correlation between user reviews and playtime is very low or non-existent. Another study by Baumann et al. (2018) uses cluster analysis in order to focus on identifying different behavioral subtypes of hardcore players.

Demand analysis and estimation. Choi et al. (2017) use reduced-form evidence (OLS, GLS) to study effects of a discount promotion, magnitude of a discount (in relative as well as in absolute value), and quantity of competitors offering price discounts to find that all but the latter have positive and significant effect on volumes of sales of Steam games. In another similar paper Choi et al. (2018) run a random effects regression on a 4-month panel data to find significant effects of various game characteristics (*e.g.* product popularity, price, user engagement) on videogame sales. Deaton (2018) uses half a year of weekly sales data in his BLP-style analysis (Berry et al., 1995) to show that both user and expert (Metacritic) reviews affect demand for videogames. Clayton (2018) studies the effect of peer influence on demand for videogames and shows that it has significant effect as well as other non-network related factors. The author's dataset has an impressive two-year span, but comes at a trade-off of collecting information on a weekly basis with no country-level price heterogeneity, which may have affected the identification of purchase prices.

Text mining analysis. Some studies examine differences between popular and unpopular Steam games using store tags (Ahn et al., 2017) and product features such as price, genres, minimum system requirements (Song et al., 2016).

1.3 Dataset and Descriptive Statistics

The main part of the dataset comes from Steam online gaming platform. I use Python to write a code that collects the data through either Steam API, Steam Storefront API, or by directly visiting Steam Store webpages and parsing their content. The code is scheduled to start running every day at midnight. As a result, I accumulate over 9 months of data spanning from June 6, 2018 to March 18, 2019. The dataset includes general information about games, prices, players' profiles, daily playing statistics, and daily exchange rates.

General information about games is collected on a weekly basis and includes the characteristics such as the name of a game, publishers, developers, release date, supported platforms, categories, genres, *et cetera.* All non-game type applications (for example, soundtracks, bonus downloadable content, movies) are excluded from the analysis. Game prices are tracked daily for all games offered in the market. The prices are available in 40 currencies that are currently supported by Steam and include a base price, a discount, and a final price. Final price is a price of the game at the moment; it is basically a base price less a discount by a base price. Customers whose local currencies are not supported by Steam pay a standard price in US dollars. In order to convert prices in local currencies to US dollars I also collect historical end-of-day exchange rates from `openexchangerates.org` with the website's API.

Individuals' profile information includes a nickname, country (self-reported), time when the profile was created, and the latest logoff date. Unfortunately, no other demographic information is available (for example, age, gender, or income level). I track playing statistics on a daily basis in order to accurately evaluate daily playtime and purchase prices. The playing statistics include identification numbers of all games currently owned by a player as well as cumulative playtime in minutes for each game. Consequently, daily playtime is defined by the difference in cumulative playtimes for every two adjacent dates. By definition, an "unopened game" is the one that has zero minutes of recorded playtime. I identify the purchase date of a game as the date when the game first appears in one's game library. Given the location of each player one can link each purchase to the corresponding price. I observe 1.4 million purchases over the course of 9 months.

One of the limitations of the dataset is that it is collected from Steam profiles with open game libraries. As of April 11, 2018 opening a game library is not a default option. Rather, it is an active choice that an individual has to make, so there may be present selection bias. Opening a game library may be required in order to play some games or use third-party online game statistics services; it also has some network benefits if one plays the same game with friends. Therefore, on the one hand this ensures that the majority of the sample users are active players; on the other hand, if being an active player is positively correlated with being an active buyer then the estimates of the overspending behavior may be upward biased.

I start with querying a pool of 185,000 consumers. Making the panel balanced and removing individuals with more than 20 hours of playtime at any day (these are very likely to be game servers, especially ones with more that 24 hours of total daily playtime) results in a sample of approximately 105,000 individuals. Appendix Figure 1.15 shows spatial distribution of the sample consumers: they are mainly located in Europe, Asia, North America, and Australia.

According to Table 1.1, the mean number of owned and unplayed games are 98 and 42, respectively. The mean share of unopened games in consumers' libraries is 28.8%. These are the highest in the countries of North America, western Europe, and Australia (Figures 1.16, 1.17, and 1.18 show spatial distribution of the above mentioned statistics).

Only 13% of sample consumers play (or at least open once) all of the games that they own as of March 18, 2019. However, these people do not have many games: the median of owned games for this subgroup is 2, while the 75th percentile is 5 games. The remaining 87% of individuals have at least one game unopened in their game libraries. The mean daily playtime is 67 minutes (conditional on playing — 183 minutes), while the mean number of games played per day is 0.54 games (1.49 games conditional on playing). Typically, around 36% of the sample individuals play daily, and if they do — only one game is played in around 68% cases. Approximately 1–13% of the sample consumers (with a mean of 3%) shop on

Table 1.1: Owned Games. Summary Statistics

	Mean	Median
No. of games owned	98	34
No. of unplayed games owned	42	8
Share of unplayed games in a game library	28.8%	25.8%

each day; conditional on a purchase they typically buy 1 or 2 games in 78% and 11% of cases, respectively. Buyers always have higher number of unopened games (both mean and median) as well as shares of unopened games in their libraries compared to no-buyers. This implies that overspending behavior is not a result of the shortage of new games in one's game library.

Next, I show that people buy new games as a replacement for the games that they stop playing. However, since consumers overestimate their willingness to try new game titles, they keep buying more than they need and as a result accumulate a lot of unopened games in their libraries. Indeed, Figure 1.2 shows that, taking country level fixed effects into account, older consumer profiles generally have higher number of games in their libraries as well as number and share of unopened games.

Since a lot of games do not have any strictly defined duration limits, I look at how many games consumers used to play but have stopped playing by the end of the observed period (abandoned games) and how many games they buy during the same time frame. Then, if a consumer buys more new games than she stops playing while keeping roughly the same number of active games at the same time one can conclude that the consumer will be unable to try all newly purchased games before she buys even more games, which will result in systematic stockpiling of unopened product. A game is defined as abandoned when a consumer played it at least once during the observed period, but stopped playing it during the last 90 days of the sample time frame. Excluding inactive consumers, on average one buys 13.5 new games and stops playing 8.5 games. According to the binscatterplot in Figure 1.3, which internalizes country level fixed effects, there is positive correlation between how many games a consumer buys and stops playing. In fact, OLS results in Table 1.2 show that, other things being equal, one buys 1.38 new games per one abandoned game.

Figure 1.4 shows the distribution of a difference between bought and abandoned games, and Figure 1.5 — the distribution of a ratio of bought to abandoned games across the sample of consumers. The difference of zero in Figure 1.4 and the ratio of one in Figure 1.5 mean that the number of games bought equals the number of games abandoned. The bunching of consumers around zero and one in the former and the latter figures, respectively, suggests that overall consumers tend to buy new games as a replacement for the games they stop playing, and there is big mass of consumers who buy excessively many games (the bought to abandoned ratio for such individuals is above one). According to Figure 1.6, consumers with higher proportion of bought to foregone games on average own higher number of unopened

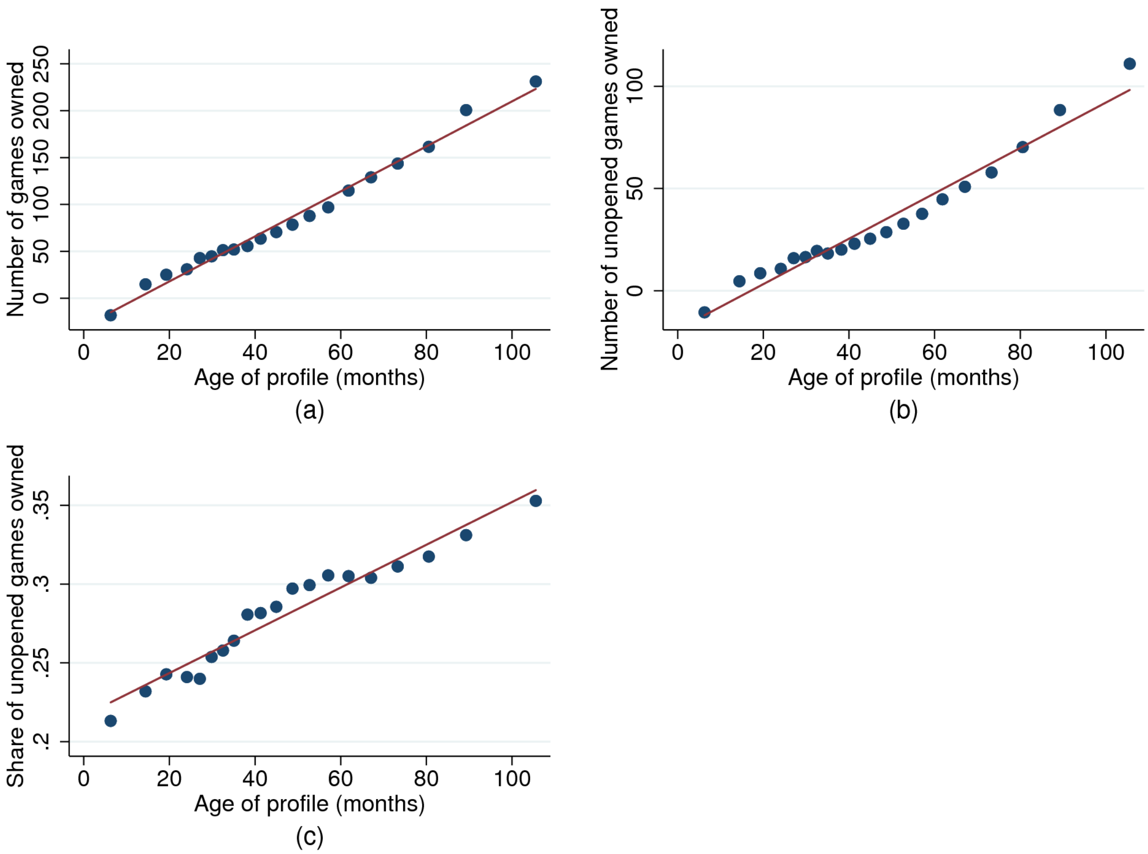


Figure 1.2: Accumulation of Games for Older Profiles

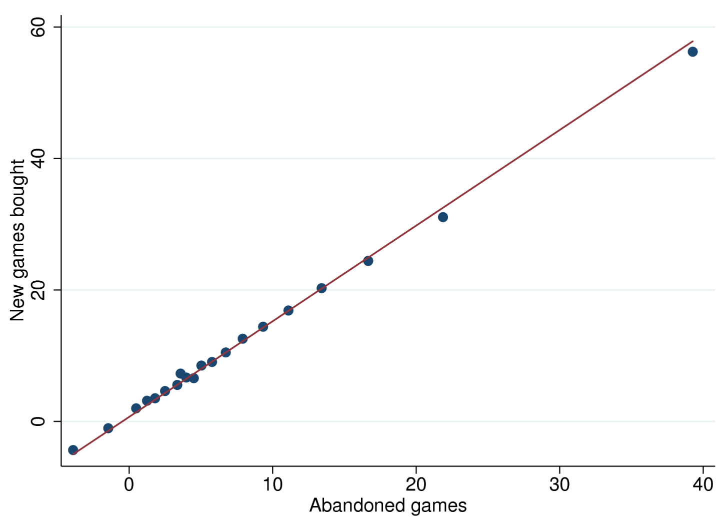


Figure 1.3: Correlation between Bought and Abandoned Games

Table 1.2: OLS Results for Number of Games Bought

	Games bought
Intercept	0.844***
	(0.085)
No. of abandoned games	1.383***
	(0.013)
No. of games owned (June 6)	0.063***
	(0.000)
No. of preexisting active games	-0.525***
	(0.012)
R^2	0.47
N	105,397

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

game titles as well as have higher shares of unplayed games in their game libraries. In graphs (a) and (b) I only analyze the observed purchases made between June 8 and December 17, 2018. I exclude the last 90 days of purchases in order to give the consumers enough time to open their games. Then I check if a purchase is still unopened as of March 18, 2019. In graphs (c) and (d) I consider all games in people's game libraries, including the games that were purchased after December 17, 2018 as well as before the start of my data collection process. All graphs take into account country level fixed effects. Visually, there is no difference in consumption patterns for the first and the second pair of graphs in Figure 1.6.

Next, one needs to check if consumers who buy more games than abandon play the same number of distinct games on a weekly basis. For each equally spaced bin of the previous two figures I run OLS regressions with a time trend and weekly number of unique games played as a dependent variable. Unfortunately, it is not possible to account for seasonality due to the data limitations. Figures 1.7 and 1.8 present the results. In both cases the trend slopes are smaller than 0.05 in absolute value, which means that on average one will play one game more or one game less in 20 weeks, depending on the sign of the slope. Surprisingly, even those consumers who buy 10 times more games than abandon still choose to play on average around two distinct games per week. Thus, the ones who overspend are not in the process of convergence to a steady state in which they can play more variety of games. Adding individual fixed effects does not change the graphs significantly. Therefore, it is natural to conclude that overspending results from heterogeneous biased beliefs of individuals with respect to the probability of wanting to try new games: one overbuys new games while keeping the number of distinct games that she plays weekly fairly constant.

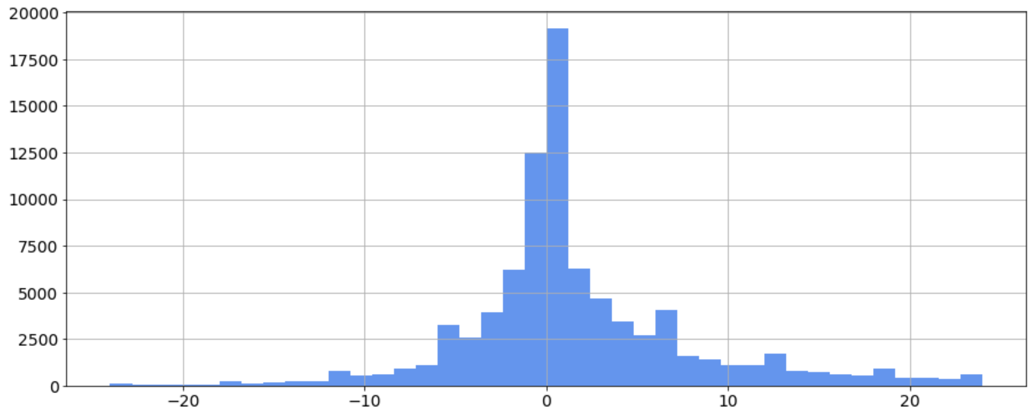


Figure 1.4: Distribution of the Difference between Bought and Abandoned Games

Note: the histogram is conditioned on buying or abandoning positive number of games and covers 87.5% of sample consumers.

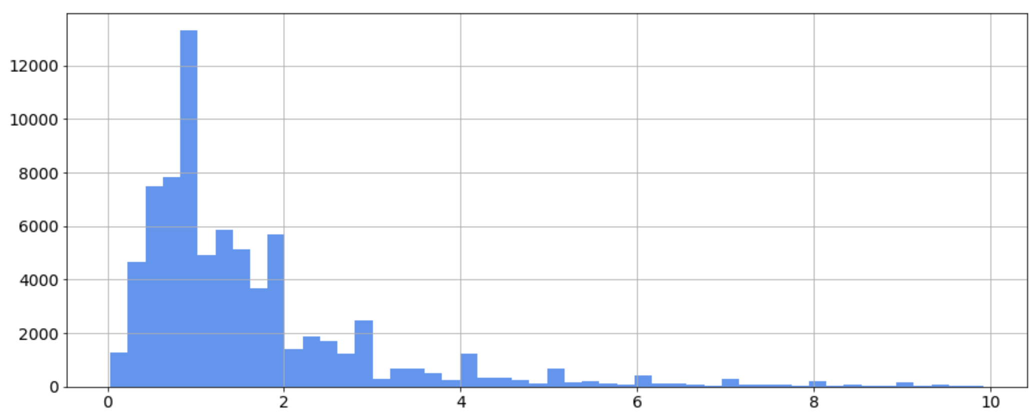


Figure 1.5: Distribution of the Ratio of Bought to Abandoned Games

Note: the histogram is conditioned on buying and abandoning positive number of games and covers 73.7% of the sample consumers. Excluded consumers: ones who (1) bought some games but did not abandon any (5.6%), (2) abandoned some games but did not buy any (8.3%), (3) did not buy and did not abandon any games (12.5%, including 6.1% of inactive people).

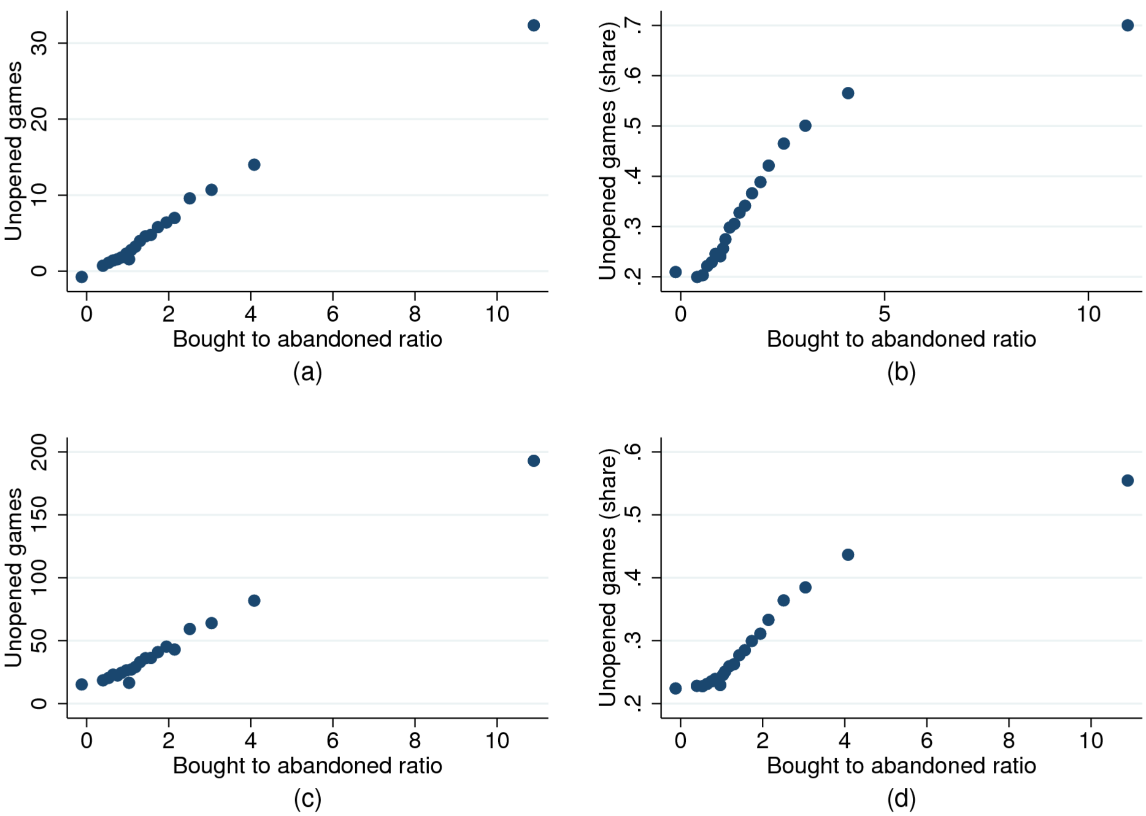


Figure 1.6: Correlation between Unopened Games and the Ratio of Game Replacement

Note: Graphs (a) and (b) only consider observed purchases from June 8 to December 17, 2018, giving the consumers at least 90 days to open their purchases. Graphs (c) and (d) take into account all games in consumers' game libraries regardless of the date of purchase, including games that were added to game libraries before June 8, 2018.

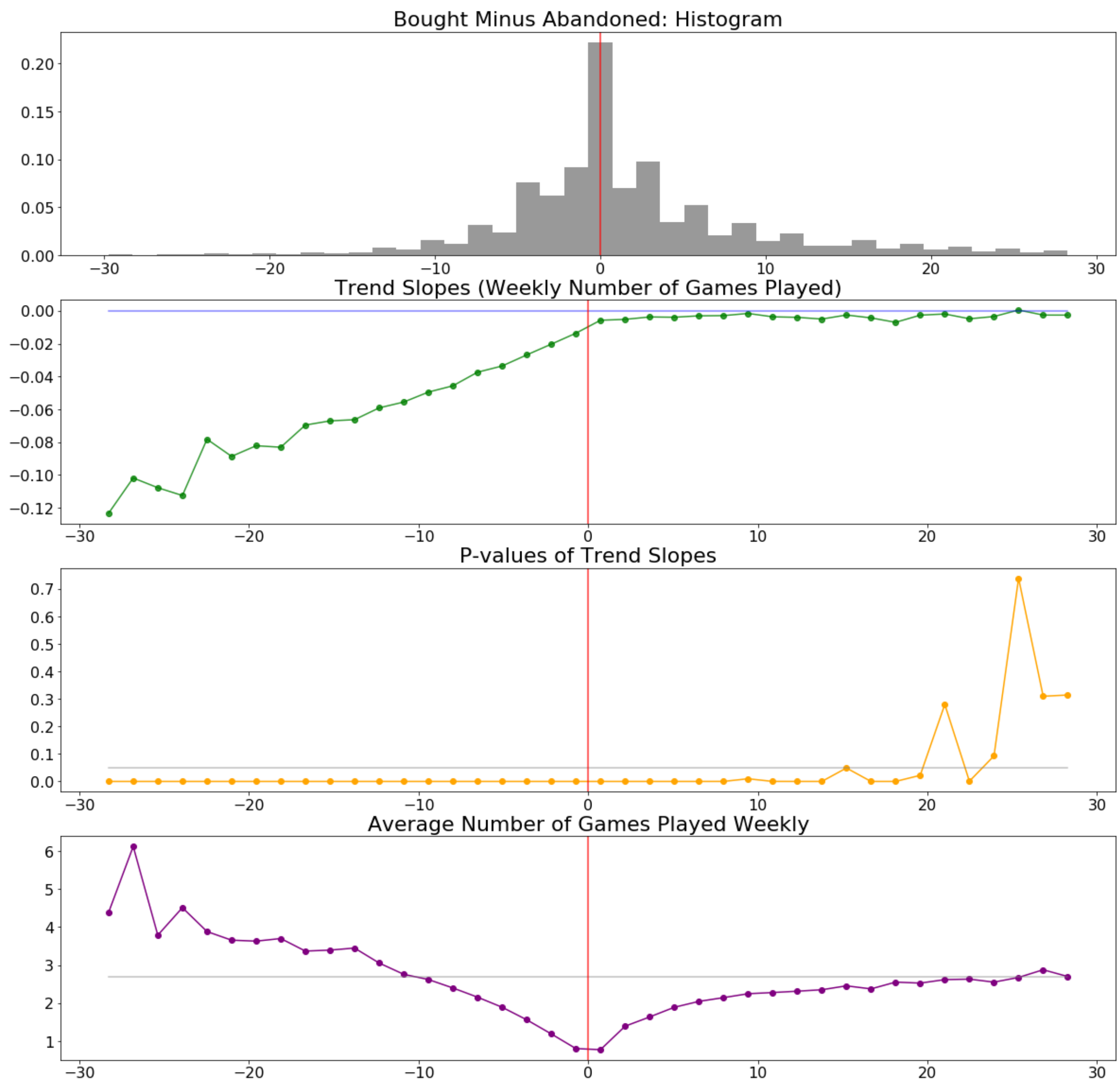


Figure 1.7: Analysis of a Weekly Number of Games Played – Part 1

Figure 1.8: Analysis of a Weekly Number of Games Played – Part 2

1.4 Mechanisms Overview

In this section I study other mechanisms that could be causing overspending behavior and show that they are not supported by the data.

Survival Analysis

The first possible explanation of overspending is that it is typical for consumers to buy games and open them later. As a result, all unopened games seem to be unopened due to censored data. However, normally a new game is either opened immediately, or not opened at all. Out of 1.4 million of observed purchases only 47% are opened. If a consumer decides to open a game she does so immediately in 60% of cases, within one day in 75% cases, and within two days in 79% of cases. The mean time to open a game is 7.3 days. Additionally, if a game is owned and unopened as of June 6 (beginning of the period) with 97.6% probability it stays unopened as of the end date, March 18.

Figure 1.9 shows survival, cumulative hazard, and hazard curves, respectively. Here a "death" event is defined by the time when a game is opened by the consumer. According to the figure, the survival function is very steep at the beginning and flattens very quickly. The cumulative hazard function is steep at first as well and concave, which implies high hazard during the first couple of weeks and little to no hazard afterwards. The smoothed hazard function asymptotically approaches zero, which supports the fact that consumers either open their games soon after purchasing them, or do not open the games at all.

Free Games

Another possible explanation is that all unopened games are free to play. According to the purchase and ownership data, only 9.7% of owned unopened games are free. Also, out of newly purchased games 19.1% are free, and out of newly purchased unplayed games 15.1% are free.

Search Cost

The next alternative mechanism is high search costs: it takes a consumer a lot of time to find the games that she likes, and once such game has been found she purchases it immediately in order not to pay additional transaction cost in the future. However, Steam has a very convenient wishlist feature. The consumer can easily add an appealing game to the wishlist and purchase it later in a couple of clicks. Accessing a wishlist from the Steam application is as easy as accessing the store page. Hence, there is no difference in transaction cost between (1) buying a game now and choosing one to play from a game library later, and (2) saving a game to a wishlist and choosing to purchase it later. In addition, unlike physical goods, games are readily available whenever one has access to the Internet, so buying a game at a full price and not playing it does not make sense as the game can be bought later at any

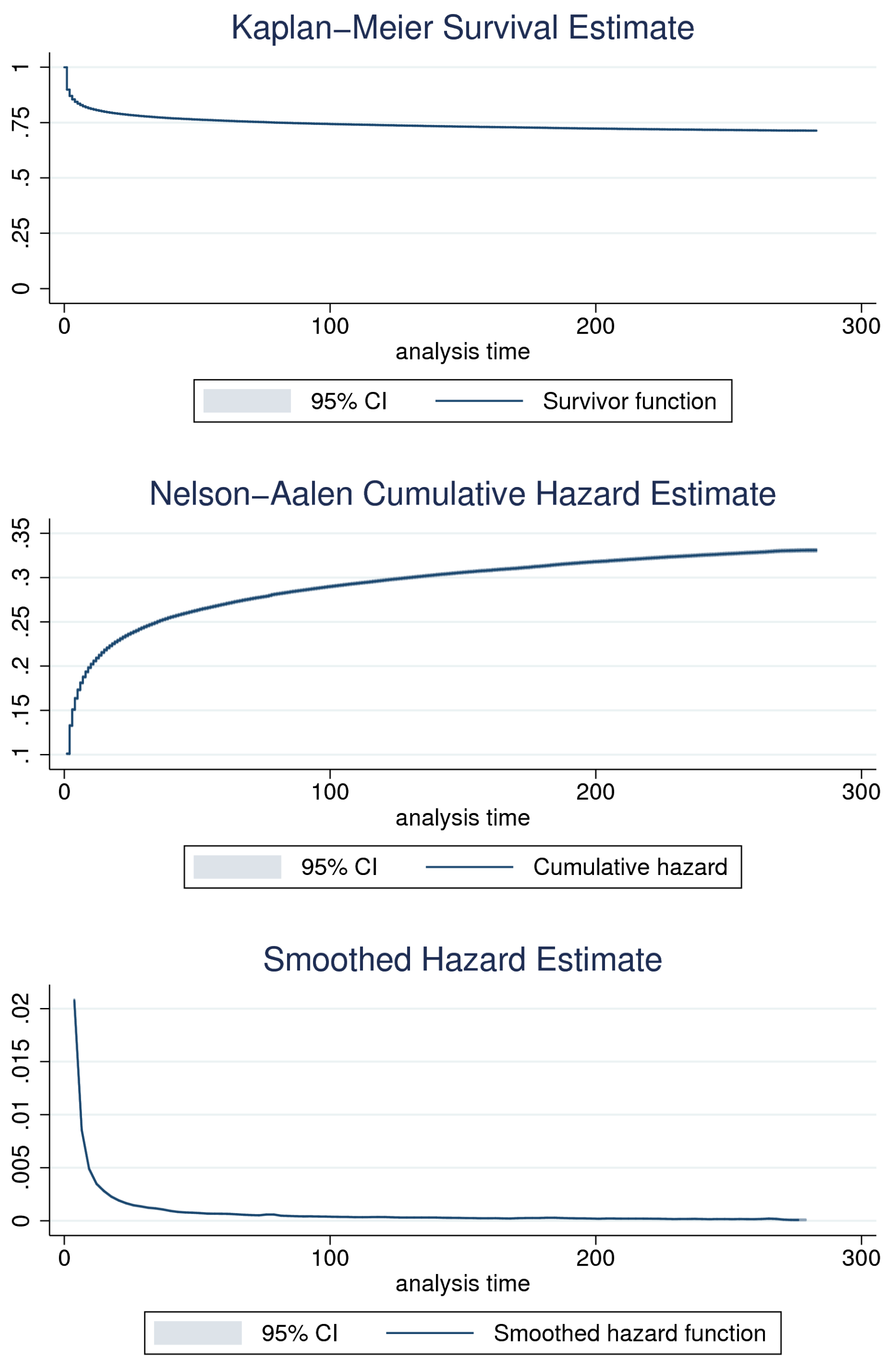


Figure 1.9: Survival Analysis

time. This behavior is also irrational in two other aspects: first, buying a game at a full price means that it may stay unopened, so it is better to postpone the decision until one is sure that she wants to play this specific game. Second, by purchasing a game at a full price now a consumer foregoes an opportunity to buy the same game at a discount later. There may be present the cost of downloading a game now versus later that might be captured by the further analysis, but given how large in size current games are it is unlikely that one will be able to download and install all games in her library on the hard drive. This physical space restriction may also partially explain why individuals play rather fixed number of games on a weekly basis. Additionally, the presence of transaction costs would make consumers buy games in bulk, but normally one buys only one or two games at a time.

Sales

Another possibility stems from the discussion above: overpurchasing may be a result of frequent sales, hence all unopened games are purchased at heavy discounts. This statement is not supported by the data either: out of 1.17 mln paid purchases 52.2% were bought on sales. 52.9% of all paid purchases are still unopened, and 56.1% of paid unopened game titles were bought at full price. Therefore, while possibly making some contribution, frequent sales are not the only reason causing excessive purchasing among consumers.

Games Heterogeneity

Next I explore heterogeneity in game characteristics. In terms of genres unopened games are typical games, not boring educational applications. The most frequent genres of unplayed games are Indie (68%), Action (44.5%), Adventure (37%), Casual (35%), Strategy (19.3%), Simulation (18%), RPG (16%), *et cetera*. According to Figure 1.10, unopened games are on average slightly cheaper then opened ones. In terms of user reviews, unopened games do not seem to be different from other games: mean and median Steam userscore of unplayed games are 72.7 and 77, while overall mean and median userscores are 72.8 and 77, respectively. Moreover, there is big overlap between played and unplayed games: only 5% of the game titles are opened by every player, and 8% of the games are never opened by anybody. The majority, 87% of games, may or may not be opened by some consumers. One could also say that all unopened games are very old games that people buy out of nostalgia. However, according to Figure 1.19, the distributions of days since release for opened and unopened games visually do not appear to be different.

Table 1.3 presents reduced form estimates with individual level fixed effects and clustered standard errors for the influence of game characteristics on consumers' playing behavior. The dependent variables are total playtime in minutes, number of days that it takes a consumer to open a game, and whether a game was opened or not. On average, price, discount percent, and total number of reviews positively affect playtime: ceteris paribus, a dollar increase in a price raises total playtime by 20 minutes, the number of days to open – by 0.007 days, and the probability of a game being opened by 0.4 percentage points. An increase in a

Table 1.3: Effect of Game Characteristics on Playing Behavior

	(1) Playtime (OLS)	(2) Days to open (OLS)	(3) Opened (Logit)
Final price (USD)	19.98***	0.00739*	0.00399***
	(0.255)	(0.00333)	(0.0000522)
Discount percent	2.388***	0.00260	0.00103***
	(0.0467)	(0.00140)	(0.0000174)
Free	-13.73*	-2.233***	0.0985***
	(5.568)	(0.124)	(0.00186)
In-app purchases	134.1***	0.779***	0.00127
	(6.844)	(0.0876)	(0.00141)
Days since release, hundreds	-6.082***	0.0782***	-0.00457***
	(0.117)	(0.00489)	(0.0000780)
Steam userscore * userscore available	-0.356***	-0.00353	0.00251***
	(0.0952)	(0.00268)	(0.0000203)
Userscore unavailable	-161.1***	-3.786***	0.0847***
	(10.01)	(0.423)	(0.00544)
Total reviews, thousands	0.754***	-0.0000883	0.000260***
	(0.0195)	(0.000178)	(0.0000102)
Games bought per purchase	-0.698**	0.501***	-0.0116***
	(0.218)	(0.0409)	(0.000709)
Constant	122.5***	5.899***	
	(8.475)	(0.251)	
Individual FE	Yes	Yes	Yes
N	1368616	677506	1356529

Standard errors in parentheses

Mean of dependent variables: total playtime - 304.62 min, days to open - 7.3 days

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

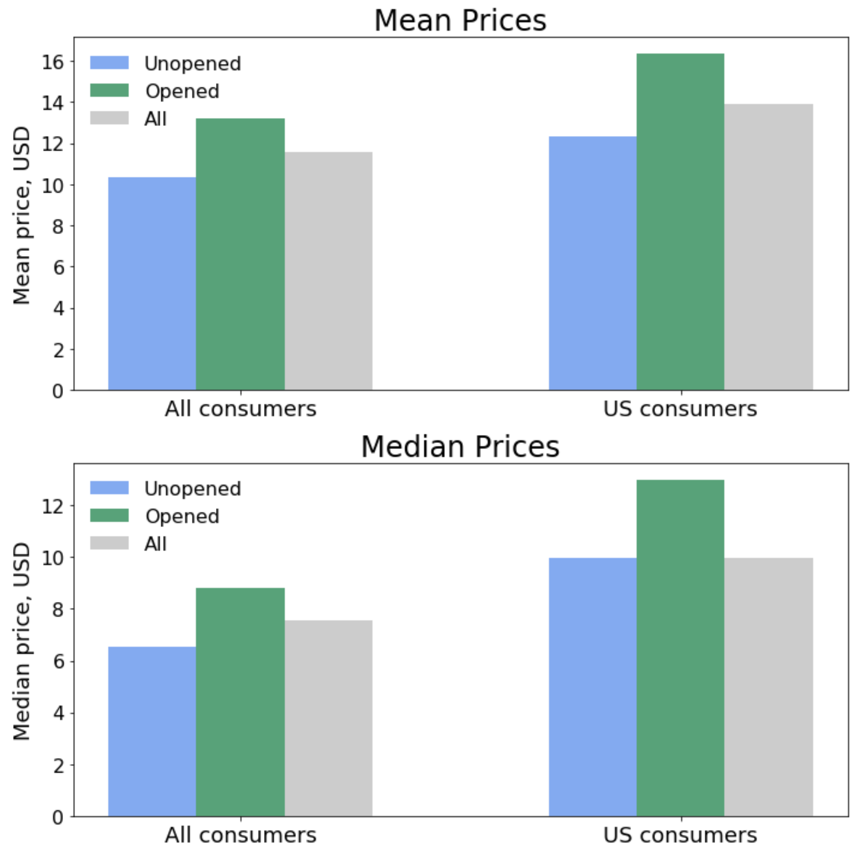


Figure 1.10: Price Comparison of Opened and Unopened Games

price discount by 10 percentage points raises expected playtime by 24 minutes, extends the number of days to open a game by 0.03 day, and increases the probability of a game being opened by 1 percentage point. The effect of peer reviews is not high in magnitude: conditional on a purchase, a thousand more reviews on the Steam Store website increases total playtime by 0.75 minutes, decreases number of days to open by 0.0001 days, and raises the likelihood of opening a game by 0.026 percentage points. Free games are played slightly less (by 14 minutes), opened faster (by 2.2 days) and are more likely to be opened (by almost 10 percentage points). Older games are also played less: a game that is older by a year is predicted to be played 22.2 minutes less, is opened slower (by 0.3 days), and it is less likely to be opened (by 1.7 percentage point). Additionally, one more game bought per purchase reduces expected playtime by 0.7 minutes, increases the expected number of days that it takes to open a game by 0.5 days, and reduces the probability of opening a game by roughly 1 percentage point. The presence of additional costs, the so-called in-app purchases, on average increases total playtime by 134 minutes, number of days to open — by 0.8 days, and probability of opening a game — by 1.3 percentage point. Few purchased game titles, however, contain in-app purchases: specifically, only 5.4% of all transactions constitute paid games with in-app purchases and 7% of all transactions are free games with in-app purchases. An increase in a userscore by 10 points is expected to decrease total playtime by 36 minutes, number of days to open a game by 0.035 days, and raises the probability of opening a game by 2.5 percentage points. Although it makes sense that better rated games are opened faster and are more likely to get opened, it is surprising to observe that they are played less at the same time. Intuitively, it may be explained by the hypothesis that due to vast variety of options to choose from the players are just overwhelmed with all the games that they own and sometimes postpone playing better games when they are time constrained in order not to spoil their impression of a better game. Absence of a userscore for a specific game title on average means lower playtime (by 161 minutes), less days to open a game (by 3.8 days), and higher likelihood of opening a game (by 8 percentage points). The last two statements, while looking controversial, make perfect sense as often new releases don't have userscores at the moment of a purchase, so this shows that consumers are more eager to open and try newer game titles.

The first regression shows that a price, whether a game is free or not, presence of in-app purchases, and absence of a userscore have the highest impact on total playtime. The effect of independent variables on the number of days it takes to open a game is quite insignificant in magnitude, except for whether a game is free or not, presence of in-app purchases, absence of a userscore, and the number of games bought per purchase. The positive impact on the odds of opening a game generally ranges within 1-4 percentage points for any of the following: $10 price increase, 10 percentage points rise in a discount, 2.74 less years since the release date, userscore higher by 10 points, 100,000 more total reviews, or 2 less games bought per purchase. A free game is more likely to be opened by almost 10 percentage points, while a game with no userscore is more likely to be opened by 8.5 percentage points.

Figures 1.11 and 1.12 graphically show average marginal effects of the regressors on the probability of opening a game. Marginal effect of one extra dollar spent on an item is positive but decreasing as the price increases and almost reaches zero for extremely expensive games. The effect of a discount percent is positive but decreasing as the total discount deepens; it stays within the range of 0.08–0.11 percentage points. As expected, negative impact of extra days since release rises in magnitude when a game is older. However, it reverses for games older than 15 years. This may be due to the fact that for really old games their age is less of a negative factor since they are considered as vintage. A change in Steam userscore has the highest impact for lower ranked games then the highest ranked ones with the 50% difference in marginal effect for extreme userscore values. A positive marginal effect of the volume of reviews is decreasing and almost reaches zero for titles with 3 million reviews. The impact of one extra game bought per purchase is always negative. The magnitude of the impact increases up to the point when a consumer reaches 20 games per purchase and then it decreases again and reaches zero for those who buy more than 100 titles at a time. Intuitively, those who buy too many games may have no real intention of opening their purchases and do so in an attempt to attract attention of other players on the platform, in particular their friends.

According to appendix Tables 1.5 and 1.6, all statistically significant coefficients from

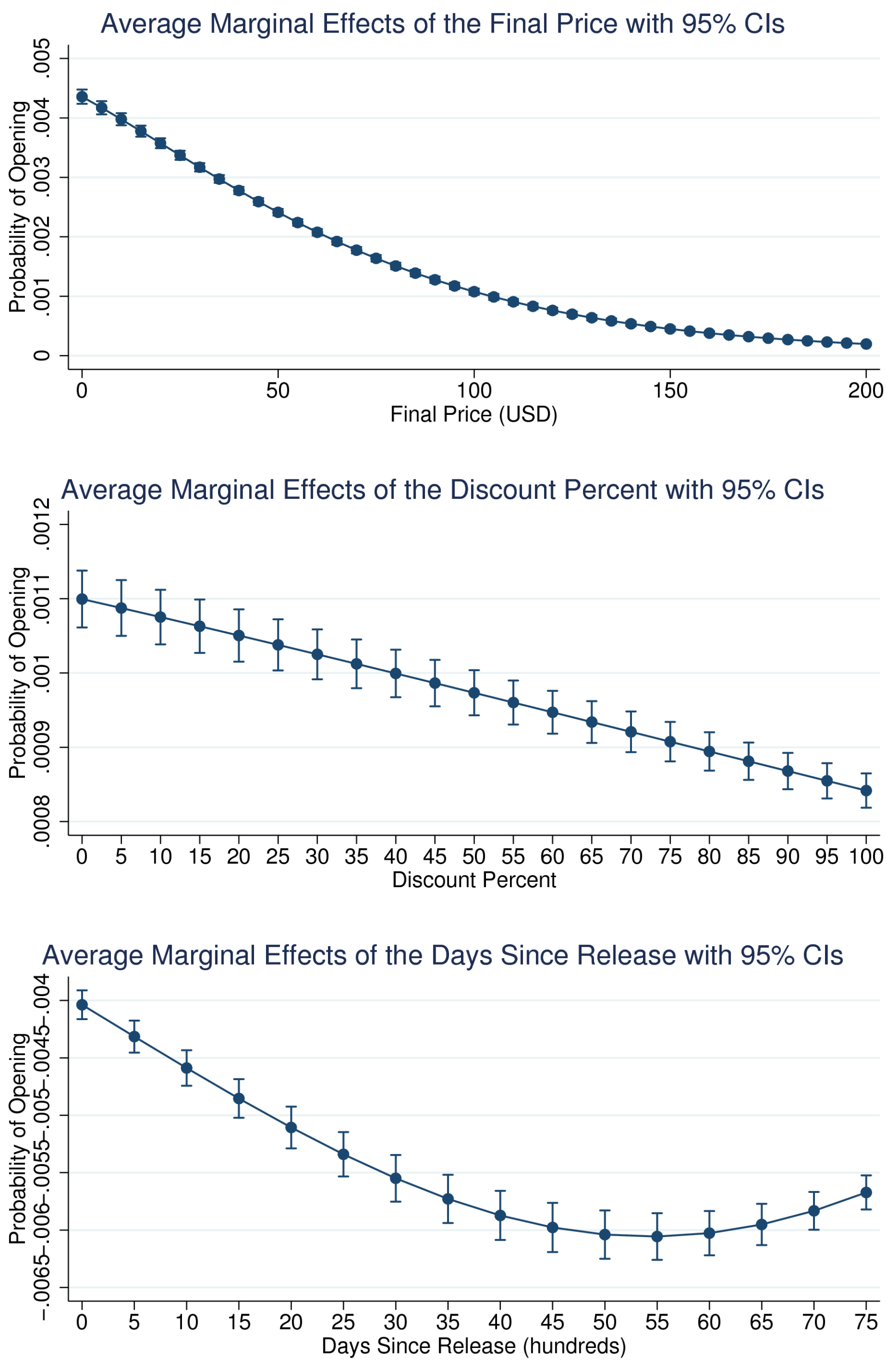


Figure 1.11: Average Marginal Effects of Price, Discount, and Days Since Release

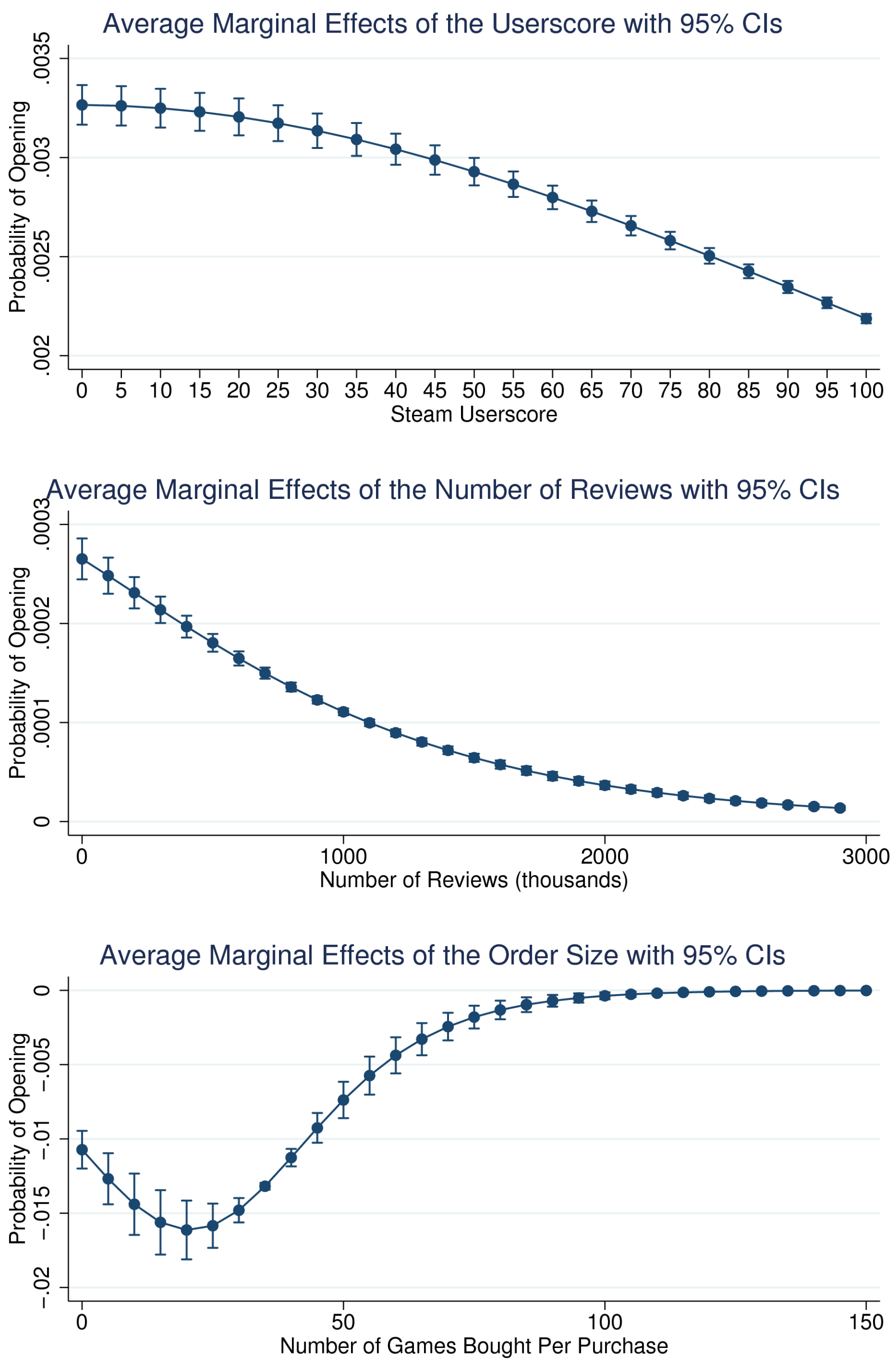


Figure 1.12: Average Marginal Effects of Userscore, Reviews, and Order Size

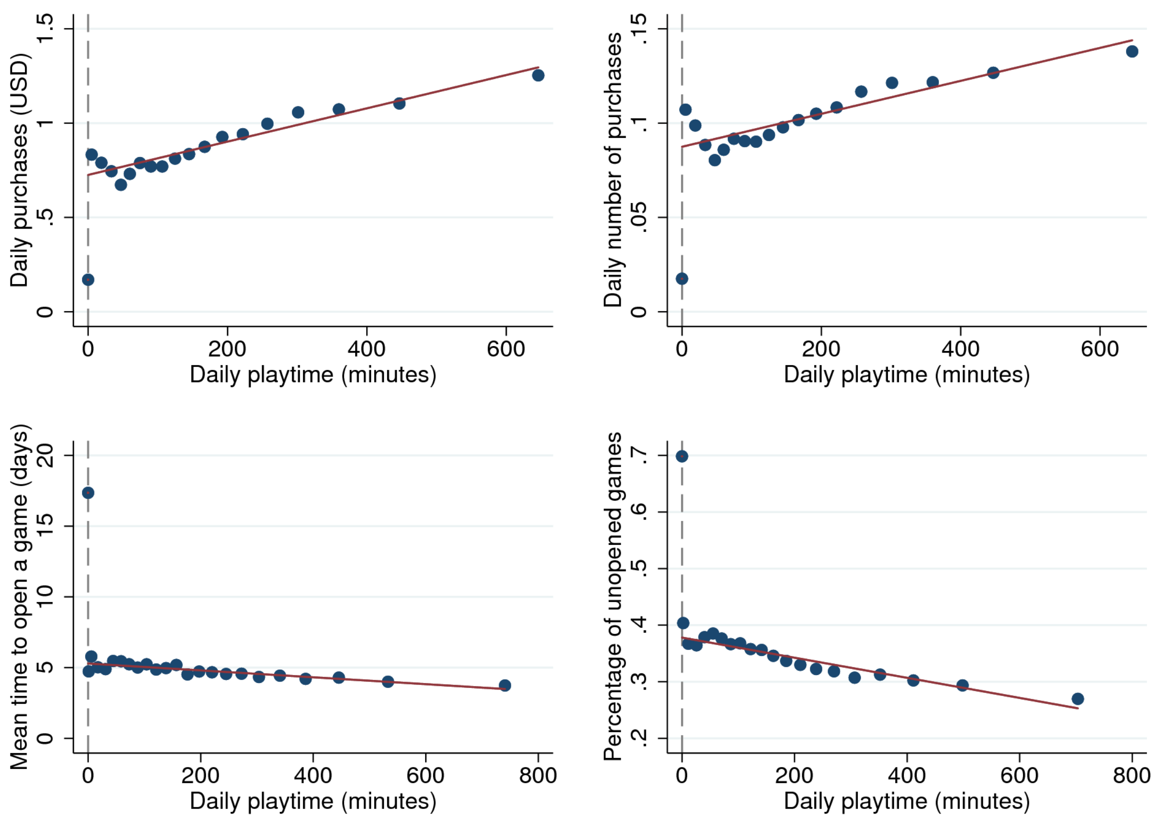


Figure 1.13: Projection Bias Check

Table 1.3 (except the one corresponding to free purchases in Table 1.5) are not sensitive to individual or country level fixed effects as well as to inclusion or exclusion from the analysis of the games that do not have any associated Steam userscore. The coefficients on price and discount percent in Table 1.6 slightly vary, but they do not appear to be significant in the most specifications. Therefore, I conclude that the magnitude of the effects of the independent variables on the probability of opening a game, while being statistically significant, is not as high in magnitude to solely explain the phenomenon and, therefore, it is heterogeneity in consumers' biases that makes individuals purchase too many games.

Projection Bias

The presence of a projection bias can be another cause of excessive purchasing: intuitively, when a consumer spends more time playing on a specific day she is likely to overestimate her future free time and eagerness to play and thus is more likely to overspend and buy more games than needed. Therefore, in the presence of a projection bias it is natural to expect positive correlation between playtime and purchases (money spent as well as quantity of items bought) on a given day. Additionally, if this is the underlying cause of the overpurchasing puzzle then the correlation between playtime and the fraction of unopened games bought on the same day should be positive as well. Figure 1.13 shows binscatterplots reflecting the aforementioned relationships. The first row of plots represents the correlation between playtime and purchases. After the initial discontinuity at zero minutes, the relation is at first negative, which suggests substitution between playtime and purchasing at low time allowances. Later it becomes positive, consistent with the theory of the projection bias. According to the second row of graphs, however, the correlation between unopened games and playtime is negative: that is, when consumers play more time on a given day it takes them on average less time to open the games that they purchased the same day. The percentage of purchased unopened games is also lower for the days when consumers play more. The latter finding goes against the theory of accumulating unopened games on the days when one plays a lot and overestimates her future preference to play more titles.

Intertemporal Substitution of Purchases

The last mechanism to explore is the possibility of the next purchase intertemporally substituting unopened games in the current purchase. If this were true then frequent purchases and long times to open games would be observed. However, the mean time to open a game is 7.3 days, while the mean observed time between purchases is 35 days. In addition, a substantial fraction of consumers (13.4%) made only one purchase during the observed period (possibly buying several games at a time). To support the evidence I run a conditional logit model with "opened" being a dependent variable. I include mean characteristics of the next purchase in the list of explanatory variables, number of days till next purchase and control for individual fixed effects. Table 1.4 presents the results. The estimates show that overall marginal effects of mean characteristics of the next purchase, while being statistically significant (except for number of reviews), are not high enough in magnitude. For example, increase in a mean discount by 10 percentage points decreases the probability of opening a game by 0.03%. If later bought games on average cost $10 more, then the currrent game is less likely to be opened by 0.13%. The marginal effects of the own characteristics of a game are consistent with the previous findings in Table 1.3. Moreover, the coefficients are the same in terms of magnitude and statistical significance and, therefore, are robust to changes of model specifications. One can also see that the number of games bought in the next purchase does not significantly affect the likelihood of opening a game in question, while the number of games bought with the game shows the same marginal effect as in Table 1.3: buying one more game decreases expected probability of opening the game in question by 1.2%.

In order to study if the next purchase crowds out the current purchase I also control for the days between two subsequent purchases. Surprisingly, waiting 100 more days to next purchase decreases the probability of opening the game by 1.3%. Figure 1.14 shows that the negative marginal effect is consistent in its sign and magnitude regardless of how many days have passed since the last purchase. The effect slightly increases in absolute value. That is,

Table 1.4: Effect of Next Purchase on the Currently Purchased Game

	(1) Opened (Logit)
Days between purchases, hundreds	-0.0132***
	(0.00226)
Games bought per purchase	-0.0115***
	(0.000724)
Games bought per purchase (next purchase)	0.000469
	(0.000294)
Free	0.106***
	(0.00213)
In-app purchases	-0.00386*
	(0.00152)
Final price (USD)	0.00427***
	(0.0000626)
Discount percent	0.00112***
	(0.0000199)
Days since release, hundreds	-0.00455***
	(0.0000891)
Total reviews, thousands	0.000303***
	(0.0000138)
Steam userscore	0.00238***
	(0.0000258)
Final price (USD, mean, next purchase)	-0.000129**
	(0.0000472)
Discount percent (mean, next purchase)	-0.0000337*
	(0.0000151)
Days since release, hundreds (mean, next purchase)	-0.000333***
	(0.0000578)
Total reviews, thousands (mean, next purchase)	-0.000000489
	(0.00000267)
Steam userscore (mean, next purchase)	0.000113***
	(0.0000331)
Individual FE	Yes
N	1252955

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

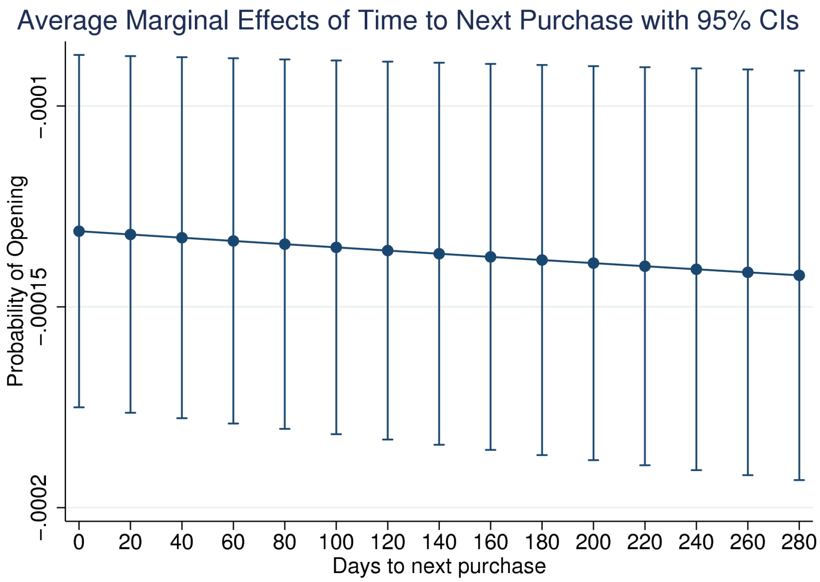


Figure 1.14: Marginal Effects of Time Between Purchases

active buyers are also more likely to try new products. The findings are not consistent with the theory of intertemporal substitution of purchases.

Therefore, it is a natural conclusion that the most likely cause of the observed overspending behavior is the individuals' biased beliefs regarding the likelihood of wanting to try a new product.

1.5 Conclusions

This paper uses a new micro-level dataset to study the phenomenon of purchasing and never trying goods in the setting of a videogames market. I explore various reasons that may be causing such behavior. I perform the survival analysis and show that normally it only takes a couple of days to open a game, and the majority of unopened games are not free to play. The cost of a repeated search is very low due to a wishlists feature, so there is no reason for consumers to purchase games right away in order to avoid future search cost. Also, overpurchasing is not driven solely by sales: around half of the paid unopened games are bought at a full price. While having a statistically significant by not high in magnitude effect, heterogeneity in game characteristics does not explain the behavior either. I check for the presence of the projection bias and show that it is not supported by the data: even though consumers are likely to buy more game titles as well as spend higher amounts of money on the days when they play most, they also open more of those games in the future. I also do not find the presence of the intertemporal substitution of purchases in the consumers' behavior.

Therefore, I argue that heterogeneity in consumer's beliefs, namely overestimation of the probability of wanting to try new games, is the most plausible explanation of the puzzle. I show that consumers buy new games as a replacement for the games that they stop buying: other things being equal, 1.38 new games are bought per one discarded game. At the same time consumers keep playing fairly constant number of games on a weekly basis, therefore they stockpile unopened games over time as they do not have enough time to open all of their purchases.

The next step in this research is to structurally model and estimate the magnitude of the bias and test if any policy changes could reduce the excessive purchasing behavior. It would be also beneficial to enrich the analysis with more demographics in order to capture consumer heterogeneity. Unfortunately, the latter is not possible to execute at the moment due to limited information about the consumers at the online gaming platform that I use for my analysis. One could also seek the alternative data sources, e.g. obtain the data from another online gaming platform which contains more demographic data.

1.6 Additional Figures and Tables

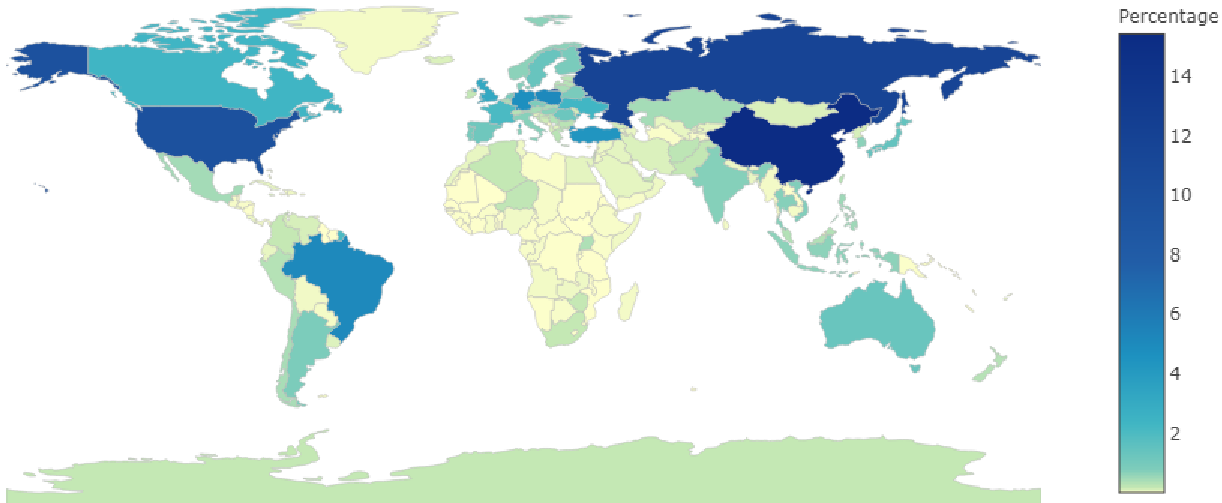


Figure 1.15: Sample Users: Geography

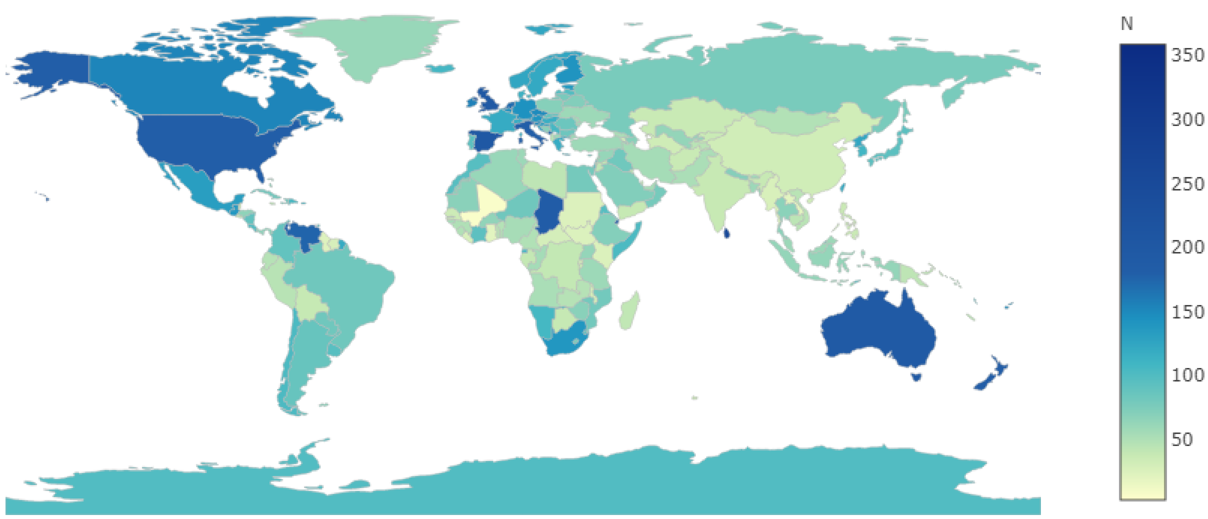


Figure 1.16: Mean of Games Owned by Country

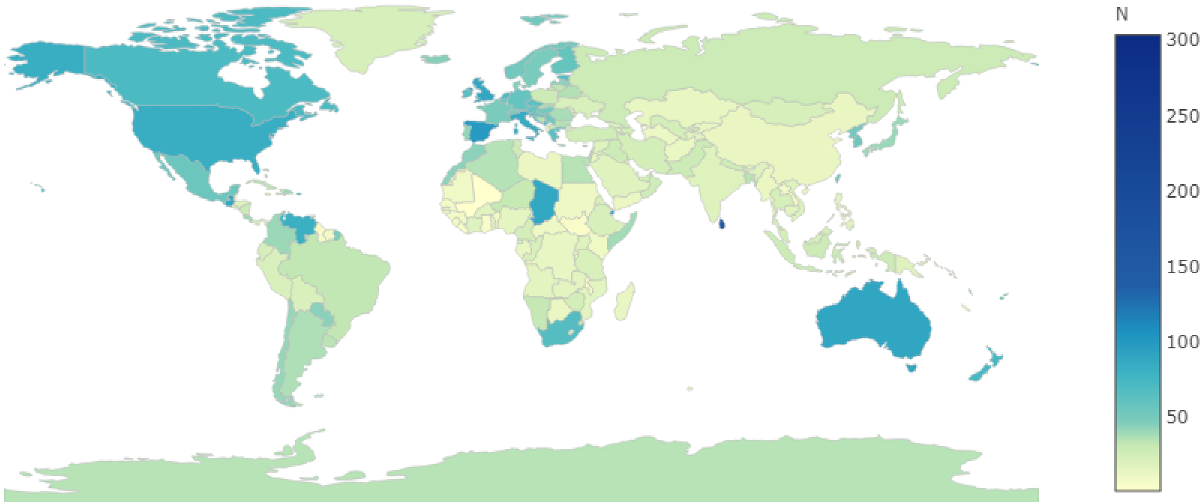


Figure 1.17: Mean of Owned Unopened Games by Country

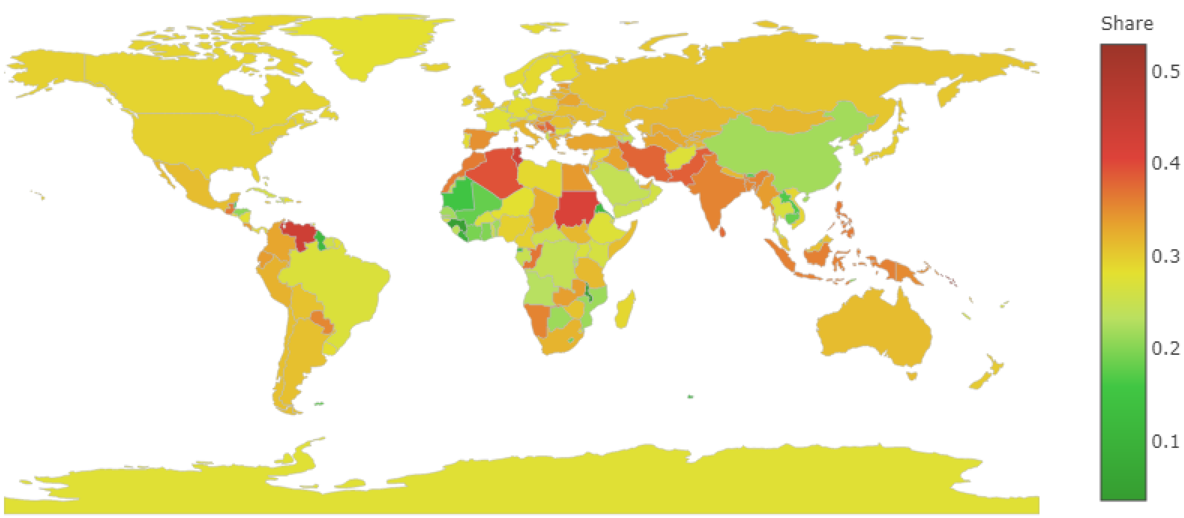
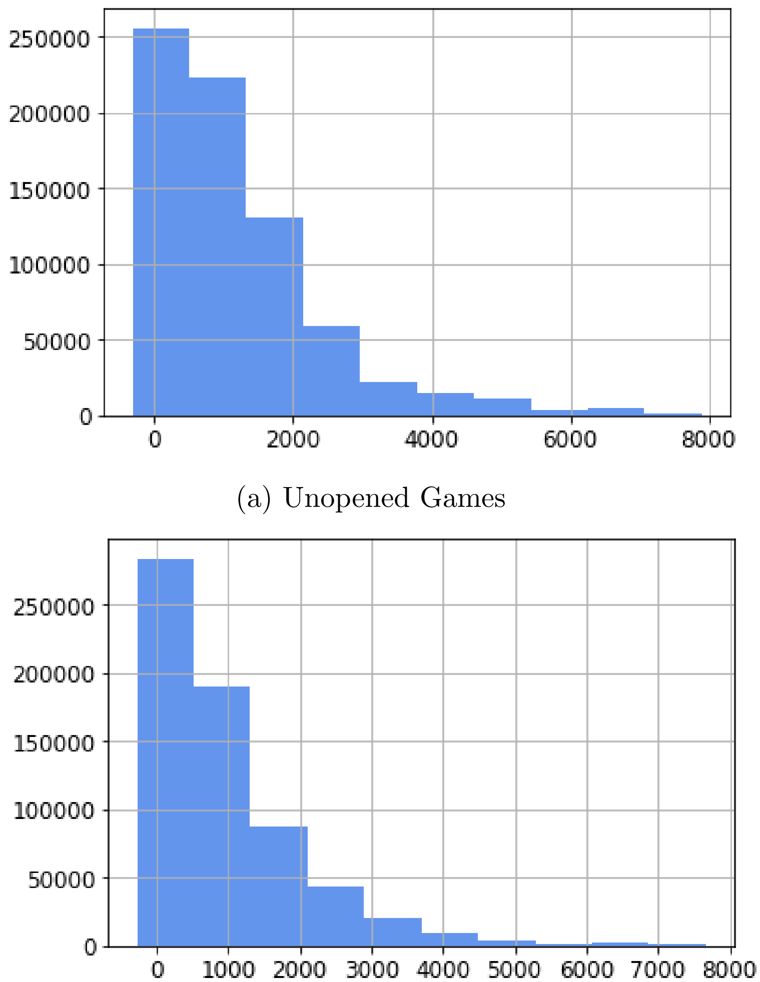


Figure 1.18: Mean of Shares of Unopened Games by Country



(a) Unopened Games

(b) Opened Games

Figure 1.19: Distribution of Days Since the Release of a Game

Table 1.5: Total Playtime – Fixed Effects Robustness Check

	(1) Playtime	(2) Playtime	(3) Playtime	(4) Playtime	(5) Playtime	(6) Playtime
Final price (USD)	20.16***	20.66***	19.98***	20.48***	21.72***	22.19***
	(0.253)	(0.262)	(0.255)	(0.264)	(2.534)	(2.676)
Discount percent	2.442***	2.497***	2.388***	2.442***	3.266***	3.327***
	(0.0461)	(0.0467)	(0.0467)	(0.0473)	(0.189)	(0.194)
Free	-16.35**	-9.950	-13.73*	-7.287	103.1***	110.2***
	(5.377)	(5.497)	(5.568)	(5.691)	(27.80)	(28.61)
In-app purchases	141.6***	140.2***	134.1***	132.6***	196.6***	195.6***
	(6.788)	(6.801)	(6.844)	(6.859)	(13.73)	(13.68)
Days since release, hundreds	-6.189***	-6.153***	-6.082***	-6.042***	-5.325***	-5.286***
	(0.114)	(0.114)	(0.117)	(0.118)	(0.334)	(0.324)
Steam userscore * userscore available	-0.415***	-0.417***	-0.356***	-0.357***	-0.444	-0.442
	(0.0956)	(0.0960)	(0.0952)	(0.0955)	(0.651)	(0.649)
Userscore unavailable	-168.3***		-161.1***		-181.0**	
	(9.978)		(10.01)		(69.38)	
Total reviews, thousands	0.709***	0.708***	0.754***	0.753***	0.747***	0.746***
	(0.0176)	(0.0176)	(0.0195)	(0.0195)	(0.0863)	(0.0865)
Games bought per purchase	-0.857***	-0.906***	-0.698**	-0.740**	-1.870***	-1.913***
	(0.252)	(0.266)	(0.218)	(0.230)	(0.302)	(0.315)
Constant	283.2***	277.4***	122.5***	116.1***	61.17	53.75
	(9.106)	(9.158)	(8.475)	(8.572)	(33.37)	(31.84)
Individual FE	No	No	Yes	Yes	No	No
Country FE	No	No	No	No	Yes	Yes
N	1368616	1352966	1368616	1352966	1368616	1352966

Standard errors in parentheses. Mean of dependent variable: 304.62 minutes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.6: Days to Open – Fixed Effects Robustness Check

	(1) Days to open	(2) Days to open	(3) Days to open	(4) Days to open	(5) Days to open	(6) Days to open
Final price (USD)	-0.0140***	-0.0119**	0.00739*	0.00984**	-0.00412	-0.00172
	(0.00365)	(0.00368)	(0.00333)	(0.00338)	(0.00449)	(0.00486)
Discount percent	-0.000759	-0.000312	0.00260	0.00302*	-0.000266	0.000206
	(0.00161)	(0.00161)	(0.00140)	(0.00140)	(0.00384)	(0.00384)
Free	-4.001***	-3.957***	-2.233***	-2.175***	-3.960***	-3.910***
	(0.142)	(0.142)	(0.124)	(0.125)	(0.361)	(0.359)
In-app purchases	0.456***	0.461***	0.779***	0.782***	0.403*	0.407*
	(0.0892)	(0.0890)	(0.0876)	(0.0878)	(0.162)	(0.162)
Days since release,	0.0697***	0.0695***	0.0782***	0.0789***	0.0659***	0.0658***
hundreds	(0.00638)	(0.00627)	(0.00489)	(0.00489)	(0.00820)	(0.00813)
Steam userscore *	0.0139***	0.0141***	-0.00353	-0.00351	0.0160*	0.0163*
userscore available	(0.00314)	(0.00314)	(0.00268)	(0.00269)	(0.00661)	(0.00658)
Userscore unavailable	-2.344***		-3.786***		-2.198***	
	(0.438)		(0.423)		(0.557)	
Total reviews,	-0.000495**	-0.000490**	-0.0000883	-0.0000822	-0.000570**	-0.000566**
thousands	(0.000166)	(0.000165)	(0.000178)	(0.000179)	(0.000199)	(0.000198)
Games bought per	0.676***	0.681***	0.501***	0.499***	0.672***	0.678***
purchase	(0.0767)	(0.0739)	(0.0409)	(0.0408)	(0.0527)	(0.0519)
Constant	4.955***	4.881***	5.899***	5.845***	4.719***	4.640***
	(0.308)	(0.304)	(0.251)	(0.252)	(0.729)	(0.724)
Individual FE	No	No	Yes	Yes	No	No
Country FE	No	No	No	No	Yes	Yes
N	677506	673284	677506	673284	677506	673284

Standard errors in parentheses. Mean of dependent variable: 7.3 days. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 2

Modeling Overspending in the Videogames Market

2.1 Introduction

In this part of the dissertation I extend the analysis performed in the previous chapter, the study of why consumers buy but do not open some of their purchases. As shown before, individuals buy games as replacements for those games that they stop playing. However, the majority of consumers consistently purchase more games than they abandon while playing fairly constant number of game titles on a weekly basis. None of the alternative mechanisms is supported by the data; therefore, I conclude that the most likely reason of excessive purchasing is the biased beliefs about the likelihood of wanting to try new products. Consistent with these findings, I build a structural model that takes the bias into account. I simulate the model and run three counterfactuals that could potentially reduce the overpurchasing behavior: debiasing consumers, decreasing variety (i.e. imposition of a lower bound on the quality of games available for purchase), and the absence of sales. The simulated results show that in the absence of sales consumers only slightly reduce their overspending. The effect of decreased variety cuts the monetary losses down as well, but the effect is not as pronounced for high userscore cutoffs. It is important to determine the cutoff that maximizes the policy effect and reduces excessive spending the most. I find debiasing consumers the most effective way to lower the unecessary expenditures, and the total effect is smaller if the individuals do not underestimate too much their probability of wanting to try a new game in the first place.

The rest of the chapter proceeds as follows. Section 2.2 briefly reviews the models used to evaluate biases and misestimated choices. Section 2.3 describes the structural model. It is a static model with the elements of dynamic behavior where every period a consumer decides how many games she wants to buy based on the number of games abandoned in the previous period and then makes a choice regarding how much time to spend on playing the games that she owns versus the outside option. Each period closes with the decision of how many games to abandon. In Section 2.4 I informally discuss identification and impose the necessary parametric and distributional assumptions in order to run the model. Section 2.5 briefly describes the simulations: I list the input parameters as well as cover the process of calibrating some of the parameters to the data that I initially scraped from the online gaming platform. Section 2.6 shows the results of the counterfactual simulations. Section 2.7 draws conclusions and discusses future work that can be done to improve the analysis.

2.2 Literature Review

Next, I briefly overview various models used in the relevant studies. This paper is broadly related to the body of literature that either deals with consumer biases, unconventional consumer decision making, or with suboptimal choices caused by irrational behavior.

For example, Hendel and Nevo (2006) measure true own and cross-price elasticities for laundry detergent taking into account that consumers tend to stockpile products during sales, thus inflating the standard estimates of the own price elasticities. The dynamic model that they use incorporates two choices: how much of a product to purchase given the individual's inventory, and how much to consume this period. DellaVigna and Malmendier (2006) study mistestimation of the future preference for gym attendance by setting up a contract choice model where consumers choose between paying a flat fee or paying per visit. When making a purchase, consumers weight the benefits against the costs, but the latter are uncertain and can be either high or low. The authors model time inconsistencies using (β, δ) discounting and also introduce partial naivete as the perceived discount $\hat{\beta}$ being in between the true value β and 1.

Another approach that is used to model biased beliefs is the introduction of different types of consumers. Oster and Morton (2005) study subscription and newsstand prices for magazines. The authors introduce a model in which consumers are divided into high and low frequency consumers who differ in how constant their willingness to pay is for magazines. One of the main findings is that due to a present bias individuals underestimate the value of magazines with delayed utility payoff but correctly estimate the value of magazines with immediate benefit and firms take this bias into account. Eliaz and Spiegler (2006) develop a theoretical model in which consumers have biased beliefs about their future demand and differ in whether they overestimate or underestimate their future demand.

It is also possible to structurally incorporate the bias into a model. Grubb (2009) studies how indivuduals choose cell phone plans. Three-part tariffs are typical for the cell phone plan market. It it a tariff that included a particular allowance in a fixed fee, but once the allowance has been exhausted the consumer has to pay steep price for additional service. Grubb argues that the reason why three-part tariffs are so widespread in this market is due to the fact that consumers are overconfident regarding their future consumption of cell phone services. He develops a pricing model similar to a screening model, in which consumers are first unaware of their demand, but learn it later on and then adjust their choices. Here overconfidence is modeled as the fact that consumers underestimate the variance of their future demand. The author considers alternative explanations of the three-part tariff, but arrives at the conclusion that overconfidence fits the observed data patterns best.

There are also papers that apply regression analysis in order to study misestimation of future consumption or benefits. Jensen (2010) uses reduced form analysis to show that it is the perceived but not real returns that define how much individuals are willing to invest in their education. Jensen conducts his own survey for school students and implements a randomly assigned intervention. He finds that that the students who were given more information on the future benefits of more education later chose to invest 0.2-0.35 more years in their education. Another reduced form paper that studies misestimation of consumption is by Bollinger et al. (2011), who examine the effect of the policy that requires chain restaurants to post calorie count next to the food options on their menus. The authors find that without calorie information consumers tend to underestimate calorie count. With the calorie information present, individuals both buy less food and substitute to less nutritious options. On average, the nutritional value of an order decreased by 6 percent after the policy implementation.

2.3 Model of Overspending

This section describes a structural model that incorporates biased beliefs of consumers. Assume a model where every day may be a normal day or a major sale day, with price distribution being different for sale and non-sale days. Consumers have static expectations about prices: although they can predict the timing of a sale, they are not aware of changes in the distribution of prices. For simplicity there is no firm side, therefore the prices are exogenous.

A brief summary of the model is as follows: a consumer i starts period t with a set of preexisting active games. She learns her purchasing shock and decides how many games to purchase (B_{it}) based on the number of games abandoned ($A_{i,t-1}$) in the previous period. Next, the consumer draws B_{it} titles from the distribution of games and decides which ones to buy. After having bought b_{it} games, some of the new purchases end up being never opened with probability γ. However, the consumer's perceived probability of not opening a game is represented by $\hat{\gamma}_i$, which is used in the purchasing decision instead of the true value γ_i. Then the consumer learns her taste shocks and decides how much time to devote to playing games. By the end of the period she decides to stop playing some of the active games.

In the more detailed description of the model that follows I intentionally drop almost all subscripts in order to keep the formulas clean and simple. Later, in the identification section, I discuss the heterogeneity of parameters across consumers as well as across countries.

When choosing how many games to buy, each individual bases her decision on how many games were abandoned in the previous period. That is,

$$B_t = kA_{t-1} + \eta_t, \tag{2.1}$$

where B_t is how many games one wants to purchase in period t, k is the proportion that shows how many more game titles a consumer wants to buy per one abandoned game (i.e. bought-to-abandoned ratio), A_{t-1} is the number of abandoned games in the previous period. η_t is a purchasing shock. When a consumer has decided on the number of purchases that she wants to make, she independently draws B_t titles from the distribution of games and buys b_t items. The individual purchases only those games which have marginal utility higher than marginal cost. She does not take into account any substitution patterns between the potential purchase and the games that she already owns. Instead, substitution of games only happens at the playing stage.

Let the taste of playing a game be θ, which depreciates depending on cumulative playtime. Basically, it represents marginal utility of playing the game one more hour:

$$\theta = \left(\overline{\theta} + \varepsilon\right) \delta^M, \tag{2.2}$$

where $\overline{\theta}$ is the baseline taste of the game, M is cumulative playtime corresponding to this game at the moment, δ is a depreciation rate, and ε is a weekly taste shock. Then, one can derive marginal utility from playing a game τ more minutes given cumulative playtime M_1:

$$\begin{aligned} U(\tau|M_1, \overline{\theta}, \delta, \varepsilon) &= \int_{M_1}^{M_1+\tau} \left(\overline{\theta} + \varepsilon\right) \delta^M dM \\ &= \left[\left(\overline{\theta} + \varepsilon\right) \frac{\delta^M}{\ln \delta}\right] \Bigg|_{M_1}^{M_1+\tau} = \frac{\left(\overline{\theta} + \varepsilon\right) \delta^{M_1}}{\ln \delta} \left(\delta^\tau - 1\right). \end{aligned} \tag{2.3}$$

The consumer stops playing a game when its taste drops below a threshold $\underline{\theta}$. Therefore, the maximum expected number of hours that a consumer can play the game given zero mean taste shock can be derived from the following equation:

$$\overline{\theta}\delta^{M^*} = \underline{\theta}. \tag{2.4}$$

Then,

$$M^* = \frac{\ln \left[\underline{\theta}/\overline{\theta}\right]}{\ln \delta}. \tag{2.5}$$

When making a purchase, an individual compares marginal cost and marginal utility of the game title. Given individual price sensitivity α, final price of a game p, base price b and a discount d ($d \in [0, 1]$), the marginal cost is:

$$\alpha p = \alpha b(1 - d). \tag{2.6}$$

The expected utility of purchasing a game depends on the perceived probability of not wanting to open a new game $\hat{\gamma}$ and the expected utility from playing a game $U^*(\cdot)$. It is the utility that one can gain from playing a game before its taste drops below $\underline{\theta}$ and it is abandoned. The utility, therefore, depends on the maximum number of minutes M^* that one can play the game given zero taste shock:

$$\begin{aligned} U^*(\overline{\theta}, \underline{\theta}, \delta) &= \int_0^{M^*} \overline{\theta}\delta^M dM = \frac{\overline{\theta}\delta^0}{\ln\delta}\left(\delta^{M^*} - 1\right) \\ &= \frac{\overline{\theta}}{\ln\delta}\left[\left(\delta^{\ln\underline{\theta}/\overline{\theta}}\right)^{\frac{1}{\ln\delta}} - 1\right]. \end{aligned} \tag{2.7}$$

Therefore, a consumer will purchase the game given that the expected marginal utility is at least as low as the marginal cost:

$$\alpha b(1-d) \leq (1-\hat{\gamma})U^*(\overline{\theta}, \underline{\theta}, \delta). \tag{2.8}$$

Every week a consumer chooses to play only those games that satisfy

$$\left(\overline{\theta} + \varepsilon\right)\delta^M > \underline{\theta}. \tag{2.9}$$

Therefore, a one-period utility from playing N games (τ_i minutes each) and choosing how much time to devote (τ_0) to the outside option with the marginal utility $\overline{\theta}_0$ can be expressed as

$$U = \sum_{i=1}^{N} \frac{\left(\overline{\theta}_i + \varepsilon_i\right)\delta_i^{M_i}}{\ln\delta_i}\left(\delta_i^{\tau_i} - 1\right) + \frac{\overline{\theta}_0}{\ln\delta_0}\left(\delta_0^{\tau_0} - 1\right) \tag{2.10}$$

subject to the restrictions

$$\sum_{i=0}^{N} \tau_i \leq T, \quad \tau_i \geq 0 \quad \forall i \in \overline{0, N}. \tag{2.11}$$

For the sake of keeping the model simple, I assume myopic consumers that only maximize their current period utility on a weekly basis. The utility for the outside option is modeled the same way as for the games except for the absence of a taste shock as well as the assumption that depreciation of the outside option does not transfer to the next day.

2.4 Identification

In this section I informally discuss the identification strategy and impose any distributional assumptions needed to identify the model in question. I am going to address the following groups of parameters that were introduced in the previous section:

- purchasing related parameters: B, k, A, b, η
- cost related parameters: α, $p(b,\, d)$
- game and taste related parameters: $\overline{\theta}$, $\underline{\theta}$, δ, M, ε, θ_0
- probabilities of not opening a game: $\hat{\gamma}$, γ

Purchasing Related Parameters

As shown earlier, the number of games that one wants to buy in the period t is defined by (2.1):

$$B_t = kA_{t-1} + \eta_t.$$

Here the bought-to-abandoned ratio k is heterogeneous across consumers. One calculates A_{t-1} as the number of games with their corresponding tastes $\overline{\theta}\delta^M$ below the threshold $\underline{\theta}$. The number of purchased games b_t is simply the quantity of the games drawn from the corresponding distribution that meet the condition (2.8). η_t is a mean zero purchasing shock:

$$\eta_t \sim N(0, \sigma_\eta^2). \tag{2.12}$$

Cost Related Parameters

In the absence of any income information, I assume α to be constant on the country level c and parametrize it as the inverse of income:

$$\alpha_c = \frac{a}{Y_c}, \tag{2.13}$$

where Y_c is mean income for each country and a is a scaling coefficient. This also implies the assumption of representative mean income consumers with no self-selection bias, which may not be true for some developing countries where access to the Internet is still considered to be a luxury. The price p for each game consists of a constant base price b and a discount d and satisfies (2.6). I identify b empirically from the pricing distribution. The discount d varies by sale and non-sale days. The joint distribution of b and d is also empirically identified:

$$Pr(b, d) = Pr(b) \cdot Pr(d|b), \tag{2.14}$$

where only distribution $f(d|b)$ changes depending on sales.

Game and Taste Related Parameters

It is not possible to separately identify individual taste parameters $\overline{\theta}$ and the lower bounds $\underline{\theta}$. However, the ratios of $\overline{\theta}/\underline{\theta}$ can be identified given the structural form imposed on the utility function. I assume that a game is abandoned when, if given a choice between playing this game and an ouside option, a consumer will prefer to spend all week on the outside option. This means that the marginal taste of a game should drop below the marginal taste of the outside option measured after the last hour of the week:

$$\overline{\theta}_i \delta_i^M < \overline{\theta}_0 \delta_0^T. \tag{2.15}$$

Therefore, I impose

$$\underline{\theta} = \overline{\theta}_0 \delta_0^T. \tag{2.16}$$

.

The rate of game depreciation δ varies across games, but does not vary across individuals. Otherwise, the pairs of (high $\overline{\theta}$, low δ) and (low $\overline{\theta}$, high δ) are observationally equivalent: it is impossible to tell whether a game was abandoned soon because it was not interesting in the first place or because it depreciated for the specific consumer too fast. Cumulative playtime M is calculated as the sum of the previous cumulative playtime corresponding to the game and the number of minutes τ one decides to devote to the game in this period. τ is identified as a utility maximizing argument. I introduce a taste shock ε in order to capture heterogeneities in playtime across the games that a consumer owns: otherwise, after some limited time marginal utility per minute for each game owned evens out and a player consumes proportional amounts of each product. A negative taste shock also explains why consumers do not play all of their active games every week. I assume the taste shock to be normally distributed with a zero mean:

$$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \tag{2.17}$$

When a consumer randomly draws a game from the pool of game titles, each draw is independently and identically distributed. With each draw she gets a game with a price p (base b and discount d), depreciation rate δ, and a baseline taste $\overline{\theta}$. I impose the following orthogonality assumptions:

$$\delta \perp\!\!\!\perp b, \quad \delta \perp\!\!\!\perp d, \quad \delta \perp\!\!\!\perp \overline{\theta}, \quad \overline{\theta} \perp\!\!\!\perp d. \tag{2.18}$$

I also assume δ to be uniformly distributed on the support $[\underline{\delta}, \overline{\delta}]$:

$$Pr(\delta) = \begin{cases} \dfrac{1}{\overline{\delta} - \underline{\delta}} & \text{if } \delta \in [\underline{\delta}, \overline{\delta}], \\ 0 & \text{otherwise,} \end{cases} \tag{2.19}$$

where

$$\underline{\delta} < \overline{\delta}, \quad \underline{\delta}, \overline{\delta} \in [0, 1].$$

Then, the probability of drawing a game with specific game characteristics $\overline{\theta}, \delta, b, d$ is:

$$Pr(\overline{\theta}, \delta, b, d) = Pr(\delta) \cdot Pr(\overline{\theta}, b, d) = Pr(\delta) \cdot Pr(\overline{\theta}|b) \cdot Pr(d|b) \cdot Pr(b). \tag{2.20}$$

The only distribution that I need to define in this expression is $f(\overline{\theta}|b)$. Since normally the price of the game reflects its production costs to some extent, I assume that conditional on drawing a game with a higher base price, the mean baseline taste is also expected to be higher. That is,

$$\overline{\theta}|b \sim N(\mu_{\overline{\theta}}(b), \sigma_{\overline{\theta}}^2), \quad \mu_{\overline{\theta}}(b) = \mu_0 + \mu_1 b. \tag{2.21}$$

I do not add any person specific components to the mean since the overall taste or distaste towards playing games will be captured by a heterogeneous taste towards the outside option θ_0, which I assume to be normally distributed:

$$\theta_0 \sim N(\mu_{\theta_0}, \sigma_{\theta_0}^2). \tag{2.22}$$

One way to make the model more versatile is to allow for heterogeneous preferences towards different game genres. Additionally, one can assume that the distribution of games that a consumer draws from is distributed similarly to the games that the consumer already owns. This will to some extent simulate recommendation services and game suggestions by the gaming platform. Unfortunately, it is impossible to know what games are recommended to each consumer on a specific day due to privacy reasons.

Probabilities of Not Opening a Game

The last group of the parameters that I discuss are the probabilities of not opening a game: a real probability γ and a perceived, biased probability $\hat{\gamma}$. Each does not vary depending on the time period, but does vary across consumers. I define the perceived probability $\hat{\gamma}$ to be higher for those games that a consumer buys on sale. I also allow for the probability of not opening a free game be different:

$$\hat{\gamma} = \hat{\gamma}_0 + \mathbb{1}\{d > 0\}\hat{\gamma}_1 + \mathbb{1}\{p = 0\}\hat{\gamma}_2. \tag{2.23}$$

As I discussed earlier in the section that overviews other possible mechanisms causing overconsumption, a consumer does not have any reason to believe that she will not open a game that is bought at a full price. Therefore, I impose

$$\hat{\gamma}_0 = 0,$$

and $\hat{\gamma}$ becomes

$$\hat{\gamma} = \mathbb{1}\{d > 0\}\hat{\gamma}_1 + \mathbb{1}\{p = 0\}\hat{\gamma}_2, \quad \hat{\gamma} \in [0, 1]. \tag{2.24}$$

Similarly, the true probability of not opening a game, γ, equals

$$\gamma = \gamma_0 + \mathbb{1}\{d > 0\}\gamma_1 + \mathbb{1}\{p = 0\}\gamma_2, \quad \gamma \in [0, 1], \tag{2.25}$$

where γ_0, γ_1, γ_2 (and hence γ) can be empirically identified.

2.5 Model Simulations

Next, I run simulations of the model described in the previous section. In order to perform the simulations, I assume specific parameter values and empirically identify them wherever possible.

First, I generate 1086 consumers with geographical locations and profile ages that are representative of the collected data. I reduce the age of each profile 4 times in order to decrease the computational burden of generating the unobservable history of decisions. For the price sensitivity α, I use income estimates from `worlddata.info` that are calculated based on the gross national product per capita published by the World Bank, IMF, and OECD (WorldData.info, 2017). I impose the following parameter values on the price sensitivity multiplier, purchasing and taste shocks, and the distribution of the marginal taste for the outside option:

$$a = 22,000,\ \sigma_\varepsilon = 10,\ \sigma_\eta = 0.75,\ \mu_{\theta_0} = 200,\ \sigma_{\theta_0} = 20,\ \delta_0 = 0.9. \tag{2.26}$$

I empirically identify γ_0, γ_1, and γ_2 for each consumer. I assume that each consumer misestimates her probability of not opening a game up to a fixed multiplier:

$$\hat{\gamma} = k_\gamma \gamma. \tag{2.27}$$

For the sake of simulations, $k_\gamma = 0.5$. I also assume that the ratio k, which defines how many games one considers to buy, is proportional to the bought-to-abandoned ratio observed in the data with the multiplier of 1.4.

In order to empirically estimate the joint distribution of a base price and a discount percent, I separately merge all observed prices for major sale days and regular days. $\underline{\theta}$ is defined by (2.16) where T is the number of hours in a week. The values imposed on the parameters that identify the distributions for depreciation rate and base taste are:

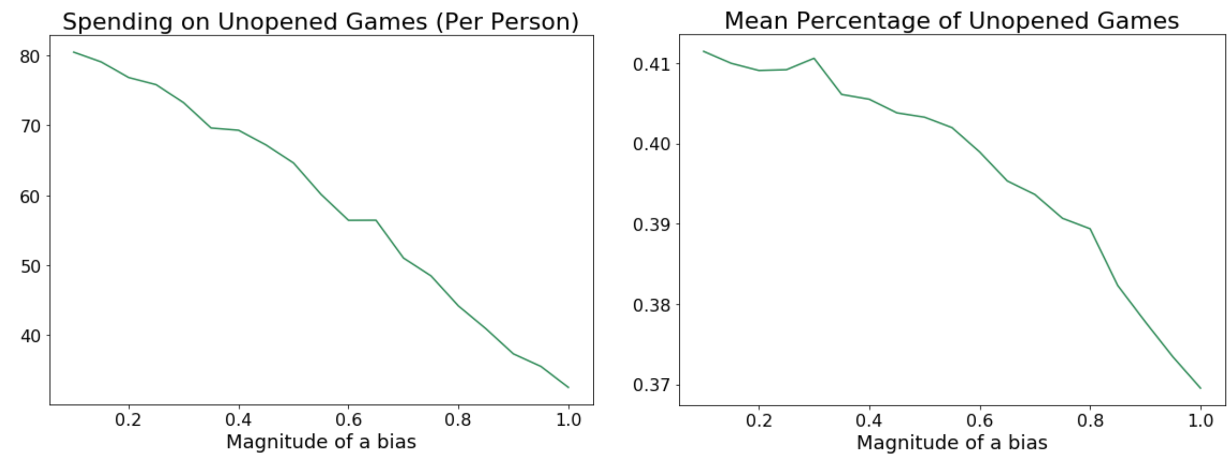
$$\underline{\delta} = 0.3,\ \overline{\delta} = 0.9,\ \mu_0 = 30,\ \mu_1 = 5,\ \sigma_{\overline{\theta}} = 20. \tag{2.28}$$

Before simulating 40 weeks of consumer decisions (the full number of weeks between June 8, 2018 and March 18, 2019), I generate previous purchasing and playing history for every individual: everyone starts with an empty game library, but acquires some game titles and plays some during the weeks that identify the age of the account. I exclude these weeks from my analysis as those are not observed in the collected data.

2.6 Counterfactuals

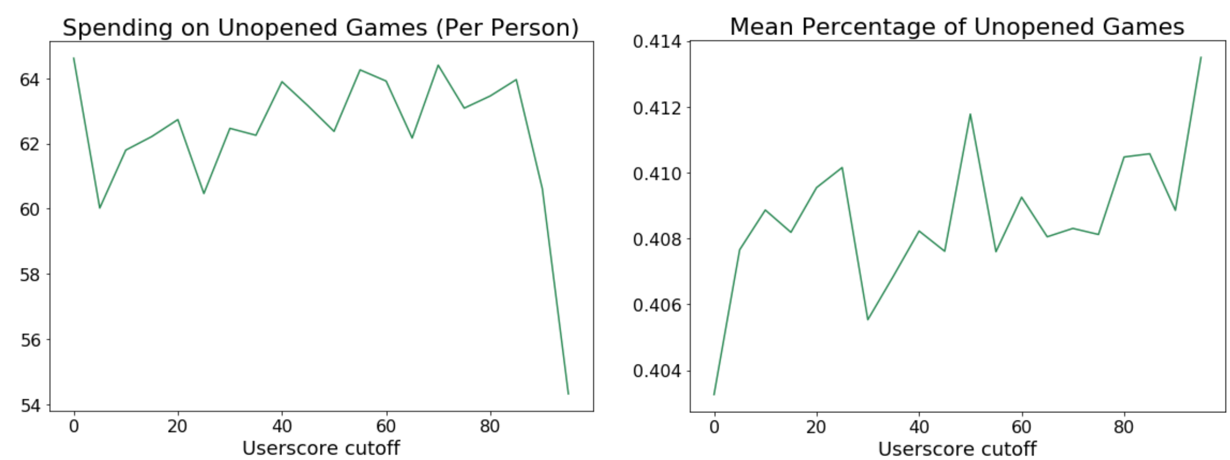
Let us discuss the counterfactuals that could potentially minimize the observed overspending behavior. I also present the simulated results for each counterfactual. Every simulation is run 10 times in order to smooth the estimates.

The first counterfactual is *debiasing consumers*. According to the model, imposing $\hat{\gamma} = \gamma$ lowers expected utility from buying a game. Therefore, this will result in consumers making less purchases and spending less money on the games that they will never open. Figure 2.1 presents the results. Figure 2.1a shows how expenditures on unopened games per capita change depending on how close the perceived probability of not opening a game $\hat{\gamma}$ is to the real probability γ. The multiplier k_γ varies from 0.1 to 1 with $k_\gamma = 1$ being the case of no consumer bias. As expected, one can see that there is a significant drop in overspending when $\hat{\gamma}$ approaches γ. That is, debiasing consumers would beneficially affect overspending, but the exact magnitude of the change depends on how close $\hat{\gamma}$ is to γ. Figure 2.1b shows that the mean percentage of unopened purchases also decreases as consumers become more aware of their bias with the maximum difference of 4 percentage points between a biased and unbiased consumer.



(a) (b)

Figure 2.1: Counterfactual Simulation — Debiasing Consumers



(a) (b)

Figure 2.2: Counterfactual Simulation — Decreasing Variety

The second counterfactual is *eliminating too much variety*. Basic models of monopolistic competition predict that sometimes total welfare may be not at the maximum if there is too much variety of products in the market. If a game does not attract any new customers (i.e. the ones who would not have bought similar games otherwise) then it is just stealing other firm's customers and does not bring any additional value. That is, efficiency-wise the game should not have been produced in the first place, which results in a welfare loss. It is possible that by restricting variety in the market, namely, imposing lower bound on game quality (userscore) one can minimize overspending behavior and achieve higher total surplus. Figure 2.2 presents the results. Figure 2.2a shows the expected change in excessive spending in the presence of a lower bound on a Steam userscore for the games available on the platform. Intuitively, cutting off the games with low score reviews will result in the change of a price and taste distribution. One is more likely to draw a better and more expensive game. On the one hand, placing a restriction on games with low userscore decreases unnecessary expenditures. On the other hand, if the lower bound is imposed too high one will be constantly drawing expensive games from the pricing distribution and some of these games will still remain unopened. That is, as the quality of purchased games raises the expected loss from not opening a game increases as well. While the unnecessary spending at first drops, the mean share of unopened purchases slightly increases compared to the benchmark, as Figure 2.2b implies. The next step in improving the results of this counterfactual is to add more game heterogeneity into the analysis and make γ dependent on other game characteristics, e.g. reviews or the age of a game.

The third counterfactual is *removing sales*. Since there is no reason to expect to not open a game that is bought at a full price, removing sales means imposing $\hat{\gamma} = 0$ for all paid games in the market. According to the results of the simulation, canceling sales slightly decreases per capita expenditures on unopened games from \$63.53 to \$61.30 per person conditional on making a purchase. The mean share of unopened games also decreases, but only very slightly (from 40.4% to 40.3%).

Therefore, the results so far show that the most effective way to reduce overspending is to make consumers aware of their bias. However, it may not be very effective if the bias is not too extreme in the first place.

2.7 Conclusions

In this paper I structurally model consumer behavior on the online gaming platform. The proposed model assumes that individuals buy new games as replacement to old game titles. It also takes into account the bias that causes some of the purchases to never be opened later. I calibrate some parameters of the model to the observed data patterns, simulate the model and run three counterfactuals: debiasing consumers, removing sales, and imposing a lower bound on the game quality. Given my input parameters, I find debiasing consumers to be the most effective way to reduce overspending behavior if the initial bias is high enough. It also decreases the percentage of unopened games in people's game libraries. Removing sales improves the status quo only slightly. Imposing a game quality floor may reduce overspending if the lower threshold is chosen optimally. However, it may also increase the percentage of unopened game titles.

Regarding the further work that one can do to improve the analysis, it will be beneficial to make the probability of not opening a game γ more heterogeneous across different game titles. It can potentially vary by other game characteristics that, although not high in magnitude, still statistically significantly affect the probability of opening a game. Adding more dynamics to the model and relaxing the assumption of a myopic consumer will also enrich the model. Finally, estimating the model so that it closely matches the observed data patterns and running the counterfactuals using the estimated parameters is the ultimate goal of this research.

Bibliography

Ahn, Sangho, Juyoung Kang, and Sangun Park (2017). What Makes the Difference between Popular Games and Unpopular Games? Analysis of Online Game Reviews from Steam Platform Using Word2vec and Bass Model. *ICIC Express Letters 11*(12), 1729–1737.

Ausubel, Lawrence M (1999). Adverse selection in the credit card market. Technical report, working paper, University of Maryland.

Balland, Pierre-Alexandre, Mathijs De Vaan, and Ron Boschma (2012). The dynamics of interfirm networks along the industry life cycle: The case of the global video game industry, 1987–2007. *Journal of Economic Geography 13*(5), 741–765.

Bauckhage, Christian, Kristian Kersting, Rafet Sifa, Christian Thurau, Anders Drachen, and Alessandro Canossa (2012). How players lose interest in playing a game: An empirical study based on distributions of total playing times. In *Computational Intelligence and Games (CIG), 2012 IEEE conference on*, pp. 139–146. IEEE.

Baumann, Florian, Dominik Emmert, Hermann Baumgartl, and Ricardo Buettner (2018). Hardcore Gamer Profiling: Results from an unsupervised learning approach to playing behavior on the Steam platform. *Procedia Computer Science 126*, 1289–1297.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen (2011). Calorie posting in chain restaurants. *American Economic Journal: Economic Policy 3*(1), 91–128.

Bounie, David, Marc Bourreau, Michel Gensollen, and Patrick Waelbroeck (2008). Do online customer reviews matter? Evidence from the video game industry.

Choi, Hoon, Dawn Medlin, and Scott Hunsinger (2017). The effects of discount pricing strategy on sales of software-as-a-service (SaaS): online video game market context. *Journal of Information Systems Applied Research 10*(1), 55.

Choi, Hoon S, Myung S Ko, Dawn Medlin, and Charlie Chen (2018). The effect of intrinsic and extrinsic quality cues of digital video games on sales: An empirical investigation. *Decision Support Systems 106*, 86–96.

Clayton, Andrew (2018). *Essays on Social Demand Estimation: Evidence from CAPS and Steam*. Ph. D. thesis, University of Washington.

Conlin, Michael, Ted O'Donoghue, and Timothy J Vogelsang (2007). Projection bias in catalog orders. *American Economic Review 97*(4), 1217–1249.

De Vaan, Mathijs, Ron Boschma, and Koen Frenken (2012). Clustering and firm performance in project-based industries: the case of the global video game industry, 1972–2007. *Journal of Economic Geography 13*(6), 965–991.

Deaton, Maxwell (2018). *Essays on Consumer Preferences Online*. Ph. D. thesis, Middle Tennessee State University.

DellaVigna, Stefano (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature 47*(2), 315–72.

DellaVigna, Stefano and Ulrike Malmendier (2006). Paying not to go to the gym. *American Economic Review 96*(3), 694–719.

Derdenger, Timothy (2014). Technological tying and the intensity of price competition: An empirical analysis of the video game industry. *Quantitative Marketing and Economics 12*(2), 127–165.

Eliaz, Kfir and Ran Spiegler (2006). Contracting with diversely naive agents. *The Review of Economic Studies 73*(3), 689–714.

Gartner (2013). Gartner Says Worldwide Video Game Market to Total $93 Billion in 2013. `https://www.gartner.com/technology/pressRoom.do?id=2614915` (accessed June 26, 2019).

Gil, Ricard and Frederic Warzynski (2014). Vertical integration, exclusivity, and game sales performance in the US video game industry. *The Journal of Law, Economics, and Organization 31*(suppl_1), i143–i168.

Gretz, Richard T (2010a). Console price and software availability in the home video game industry. *Atlantic Economic Journal 38*(1), 81–94.

Gretz, Richard T (2010b). Hardware quality vs. network size in the home video game industry. *Journal of Economic Behavior & Organization 76*(2), 168–183.

Grubb, Michael D. (2009). Selling to Overconfident Consumers. *American Economic Review 99*(5), 1770–1807.

Grubb, Michael D and Matthew Osborne (2015). Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review 105*(1), 234–71.

Hendel, Igal and Aviv Nevo (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica 74*(6), 1637–1673.

Jensen, Robert (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics 118*(4), 1209–1248.

Oster, Sharon M and Fiona M Scott Morton (2005). Behavioral biases meet the market: The case of magazine subscription prices. *The BE Journal of Economic Analysis & Policy 5*(1).

PCGamesN (2017). With $4.3 billion in sales, 2017 was Steam's biggest year yet. `https://www.pcgamesn.com/steam-revenue-2017` (accessed September 14, 2018).

Read, Daniel and Barbara Van Leeuwen (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational behavior and human decision processes 76*(2), 189–205.

Sifa, Rafet, Anders Drachen, and Christian Bauckhage (2015). Large-scale cross-game player behavior analysis on steam. *Borderlands 2*, 46–378.

Song, Sungmin, Nam-Wook Cho, and Taegu Kim (2016). Success Factors of Game Products by Using a Diffusion Model and Cluster Analysis. *Journal of Korean Institute of Industrial Engineers 42*(3), 222–230.

Venkatraman, N and Chi-Hyon Lee (2004). Preferential linkage and network evolution: A conceptual model and empirical test in the US video game sector. *Academy of Management Journal 47*(6), 876–892.

WorldData.info (2017). Average income around the world. `https://www.worlddata.info/average-income.php` (accessed May 2, 2019).

Zhu, Feng and Xiaoquan Zhang (2006). The influence of online consumer reviews on the demand for experience goods: The case of video games. *ICIS 2006 Proceedings*, 25.